Assessment of host genetics and environmental factors in shaping the gut microbiota

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Eric Juan Carlos Gálvez Bobadilla

aus Fusagasugá / Kolumbien

1. Referentin:	Professorin Dr. Petra Dersch
2. Referent:	Professor Dr. Karsten Hiller
eingereicht am:	01.10.2018
mündliche Prüfung (Disputation) am:	12.12.2018
Druckjahr	2019

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch die Mentorin der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

Gálvez EJC*, Iljazovic A, Gronow A, Flavell RA, Strowig T**. Shaping of intestinal microbiota in Nlrp6 and Rag2 deficient mice depends on community structure. Cell Rep. (2017).
* First author, **Corresponding author

Tagungsbeiträge

Eric J.C Gálvez and Till Strowig: Low Complexity Microbiota mice (LCM) A stable and defined gut microbiota to study host-microbial interactions (Oral Presentation). 7th Seeon Conference on „Microbiota, Probiota and Host“, 04-06 July 2014. Kloster Seeon, Germany.

Eric J.C Gálvez and Till Strowig: Composition and functional dynamics of novel gut commensal species of *Prevotella* (Oral Presentation). EMBO|FEBS Lecture course: The new microbiology, 24 August – 1 September 2016. Spetses, Greece.

Acknowledgments

I would like to express my deep gratitude to my mentor, Dr. Till Strowig, for his constant support during the past years. Thank you for your patient guidance, enthusiastic encouragement and all the great advice without which this thesis would have not been possible.

Special gratitude to Prof. Dr. Emmanuelle Charpentier for her support thrust and help in my research.
I also would like to thank Prof. Dr. Marc Erhardt, for his suggestions and advice to improve my PhD project.
Special acknowledge to Prof. Dr. Petra Dersch and Prof. Dr. Karsten Hiller for agreeing to be on my examination committee, as well as Prof. Dr. Anett Schallmey for serving as the head of my Ph.D. dissertation.
I am also very thankful to the HZI Graduate School for the financial support and for the several career development opportunities provided.
Exceptional acknowledgements to all the members of the MIKI group. Thank you for all the help, technical support and friendship. I am especially grateful to Aida Iljazovic, Urmi Roy, Robin Lesker, Adrian Blazèjewski and Sophie Thiemann. It has been a pleasure to work with you guys and have the privilege to have you as friends.
My gratitude goes to all my friends that I had the pleasure to meet in Braunschweig and Berlin. You made this journey an amazing experience that will remain in my mind and my heart.
Most importantly I want to thank my fiancée Lina Zárate, who has been next to me during my scientific carrier. Thank you for your constant love, support and motivation.
Finally, I wish to thank my sibling Alejandro and my parents Martha and Erickberto, who have provided me through moral and emotional support in my life.

A Valeria y Victoria

"If you are not having fun, you are not learning.
There is a pleasure in finding things out."
~ Richard Feynman

Summary

The gut microbiome harbors a complex network of bacteria, archaea, and viruses. Governed by members of the domain Bacteria, the microbiome carries out essential physiological functions, such as promoting the development of the immune system and providing protection against pathogens. A well-balanced composition is fundamental for the health and fitness of the host. In contrast, microbial imbalances have been linked with various pathologies such as intestinal inflammation and metabolic syndrome. Therefore, it is fundamental to understand the major determinants of the gut microbiota composition and stability. Yet, studies have observed that the gut microbiome is shaped by multiple components, but the role of host genetics and environmental factors remains incompletely understood.

Here, we evaluate the composition and stability of the gut microbiota in enhanced specific pathogen free (eSPF) mice, a standardized model to study gut microbiota interactions. Furthermore, we used the eSPF model to assess the relative contribution of the NLRP6 sensor and deficiencies in adaptive immunity (*Rag2*$^{-/-}$) on shaping the gut microbiota. Our results showed that the impact of NLRP6 and adaptive immunity on microbiota composition depends on community structure and particularly influences pathobionts such as Helicobacter spp., but not commensals.

Next, using omics approaches we characterized the genomic diversity and functional niche of novel species of the Prevotellaceae, an uncharacterized bacterial family which its abundance has been associated with health and host disease. Our results unveiled high genome diversity at the strain level. In addition, the functional characterization in vivo reveled an up-regulation of distinct polysaccharide utilization loci (PULs) associated with xylan degradation. Further, we identified that a specific repertoire of PULs may enable the strain P. intestinalis to outcompete in mice fed with complex plant polysaccharides. Our study model dissects the specialized metabolic niche of *Prevotella* and uncovers potential genetic determinants for their abundance and prevalence in the gut.

Collectively, these data demonstrate how host genetics and environmental factors shape the composition and function of the gut microbiota. With the aim to translate the understanding of the microbiome gained from our model, further studies using genetically engineered *Prevotella* strains are required to determine the molecular bases for colonization and abundance.

Zusammenfassung

Unser Darmmikrobiom besteht aus Bakterien, Archaeen, Eukaryoten und Viren. Darmbakterien führen essentielle physiologische Funktionen aus, wie Stoffwechselvorgänge, die Modulation des Immunsystems und dem Schutz vor Krankheitserregern. Mikrobielle Ungleichgewichte werden mit unterschiedlichsten Erkrankungen in Verbindung gebracht. Um die Rolle des Darmmikrobioms in Gesundheit und Erkrankung zu verstehen, ist es von grundlegender Bedeutung, dessen Zusammensetzung zu analysieren, sowie den Zusammenhang mit Wirtsgenetik und Umweltfaktoren zu verstehen.

Um den Einfluss verschiedener Faktoren auf das Darmmikrobiom zu analysieren, wurde die Zusammensetzung und Stabilität der Darmmikrobiota in spezifischen Pathogen-freien (eSPF) Mäusen untersucht, einem stabilen und reproduzierbaren *in vivo* Modell zur Untersuchung von Darm-Mikrobiota-Interaktionen. Weiterhin wurde dieses Modell genutzt, um den Einfluss des NLRP6-Sensors und des Fehlens der adaptiven Immunität auf die Darmmikrobiota zu untersuchen. Unsere Ergebnisse konnten zeigen, dass dies abhängig ist von der bakteriellen Gemeinschaftsstruktur, und insbesondere dem Vorhandensein von Pathobionten wie *Helicobacter spp.,* abhängt, jedoch nicht von Kommensalen.

Außerdem charakterisierten wir mithilfe multipler "Omics"-Ansätze die genomische Diversität und funktionelle Nische neuartiger Spezies der Prevotellaceae, einer abundanten Familie kommensaler Darmbakterien, deren Vorkommen mit Gesundheit und Erkrankungen assoziiert wurde. Unsere Ergebnisse identifizierten eine hohe Genomvielfalt in der Gattung *Prevotella*, sowie eine Hochregulierung von verschiedenen Polysaccharidverwendungsorten (PULs), die mit dem Xylanabbau verbunden sind. Weiterhin konnten wir zeigen, dass ein spezifisches Repertoire an PULs dem Stamm *Prevotella intestinalis* ermöglicht, im Darm von Mäusen mit ballaststoffreicher Ernährung andere Stämme zu dominieren. Unser Studienmodell entschlüsselt die metabolische Nische der *Prevotella*-Arten und deckt mögliche genetische Determinanten für deren intestinale Prävalenz auf.

Zusammenfassend zeigen diese Daten, wie Wirtsgenetik und Umweltfaktoren Zusammensetzung und Funktion der Darmmikrobiota beeinflussen. Weitere Studien sind erforderlich, um das aus unserem Modell gewonnene Verständnis des Mikrobioms anzuwenden und die molekularen Mechanismen zu verstehen, welche zuständig sind für Kolonisation und Prävalenz kommensaler Bakterien im Darm.

Table of content

Abbreviations

ASF	altered Schaedler flora
ATP	adenosine triphosphate
CAZymes	Carbohydrate-Active Enzymes
CAZyome	total profile of Carbohydrate-Active Enzymes
CD	Crohn's disease
Cec	cecum
DC	distal colon
DNA	deoxyribonucleic acid
DSS	dextran sulfate sodium
EDTA	ethylenediaminetetraacetic acid
eSPF	enhanced specific pathogen free
FT	fecal transplantation
GF	germ-free
GI	gastrointestinal
GWAS	genome-wide association study
HZI	Helmholtz Centre for Infection Research
IBD	inflammatory bowel disease
IFN	interferon
IgA	immunoglobulin A
IL	interleukin
LDA	linear discriminant analysis
LPS	lipopolysaccharide
MAMPs	microbe-associated molecular patterns
mGWAS	microbiome genome-wide association study
MLN	mesenteric lymph nodes
MyD88	myeloid differentiation primary response protein 88
NCI	National Cancer Institute
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	next generation sequencing
NLRP	NOD like receptor protein
NLRs	nucleotide-binding domain and leucine-rich repeat-containing receptors
NMDS	nonmetric multidimensional scaling
NOD	nucleotide-binding oligomerization domain
NOD2	nucleotide oligomerization domain 2
PBS	phosphate-buffered saline
PC	proximal colon
PC	proximal colon
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
PRRs	pattern recognition receptors
PSA	polysaccharide A

PUL	polysaccharide utilization loci
RAG	recombination activating gene
RNA	ribonucleic acid
ROS	reactive oxygen species
SCFA	short-chain fatty acid
SCFAs	short-chain fatty acids
SDS	sodium dodecyl sulfate
SFB	segmented filamentous bacteria
SI	small intestine
SNP	single nucleotide polymorphism
SPF	specific pathogen free
TLR	toll like receptor
TLRs	toll-like receptors
Treg	regulatory T cell
UC	ulcerative colitis
WT	wild type

List of figures

1 Introduction

1.1 General Introduction

If we look up the sky at night, we may contemplate about 2 trillion of galaxies (*1*) and if each of these galaxies contains an average of 100 billions of stars (Milky Way's estimate), we could look up the tremendous estimate of 10^{24} stars in the universe. If we then glance down again and realize that individuals can harbor on and in their bodies a mere of 10^{13} microbes, we could picture that the study of the microbiome "is a story of big numbers, complex interactions and multidimensional networks" (*2*). Over the last decades, astrophysicists have developed theoretical models to understand the universe using observational data from powerful radio telescopes. In the meantime, in the microbiome science we are still collecting data with the aim to identify "who are there" and "what are they doing", while our understanding of the rules that govern host-microbial interactions remain relatively limited. However, the synergy of significant advances between physics, chemistry and biology has allowed developing better tools to massively explore the microbiome networks. Analogous to radio telescopes, different "omics" technologies are the equivalent tools in the microbiome field, which continuously are providing insights into the genetics, transcriptional and metabolic phenotypes that in the future could be integrated and converted in theoretical models. This analogy from Macpherson and McCoy contrasting the microbiome and the universe has been on the most simples and beautiful concepts that I have found to simplify the scale and the challenges in the microbiome scenario.

1.2 The Gut microbiome

The microbiome encompasses diverse populations of bacteria, viruses, archaea, and eukaryotes that populate many body sites of multicellular hosts (*3*). Recently, it has been estimated that we, humans, can host the same number of bacteria in our body as the number of own cells (*4*). From this vast bacterial biomass, up to 98% is located in the gastrointestinal tract (GI) (*5*). The gut itself is a complex system, which supports distinct microbial ecosystems across different anatomical sites. Notably, the highest diversity of bacterial strains is found in the large intestine (*6*). The large intestine, also known as the large bowel or colon is a fascinating environment; on one side, the lumen hosts a dense mix of strictly anaerobic bacteria, whilst the lamina propria harbors dynamic populations of aerobic immune cells, which are separated by a tiny layer called the intestinal epithelium (*7*). In the lumen, the microbiome encodes up to 9.8 millions of genes (*8*), which it turns out in 500-fold more foreign genes in contrast with the genes that are encoded in the human genome. As an example, the carbohydrate-active enzymes (CAZymes) repertoire encoded by human genome is minimal (17 enzymes approx.) (*9*) compared to the immense carbohydrate degradation machinery encoded in the genome of *Bacteroides thetaiotaomicron* (261 GHs enzymes approx.) (*10*), a common commensal bacterium in the human microbiome.

Of the dense bacterial biomass, the human gut microbiota can be composed by over a 1000 potential different species or operational taxonomic units (*11*). Interestingly, among the 114 bacterial phyla described until now on the earth, the vast diversity of the gut microbiome is mainly from 2 phyla, Firmicutes and Bacteroidetes (*12*). According to the capabilities of bacteria to colonize the GI, they can be transient or permanent (*13*).

Transient bacteria are the microbes that originate from the environment and do not permanently establish themselves in the intestinal tract due to the lack of the gene pool for adaptation, or the inability to compete with the established microbiota. In contrast, permanent bacteria are long-term members of the microbial community and depending on their functionality and the interplay with the host they can be grouped as commensals, symbionts or pathobionts (see definitions in Figure 1.1).

During the last decade, studies in germ-free (GF) animals and the application of culture-independent methods have established the importance of the microbiota for host physiological functions in health and disease (*14*). Among the health-associated phenotypes or symbiosis, researchers have focused on investigating the role of the microbiome in proper immune system development, nutrient metabolism and protection against infection. On the other hand, also adversary effects associated with the microbiome have been characterized, namely contributions to metabolic syndrome, inflammation, autoimmune disease and cancer among the others (*15*).

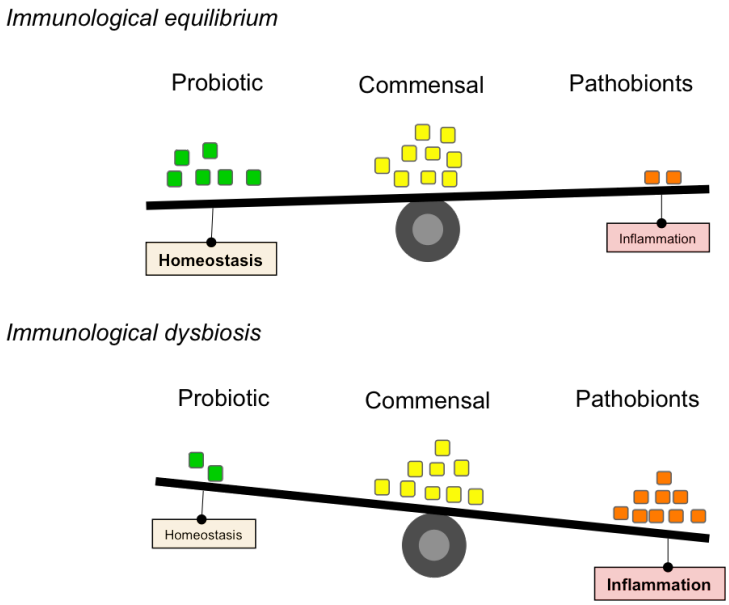

Figure 1.1: Host immunological equilibrium and dysbiosis according to the composition and functional role of the microbiome.

Probiotic (symbiont): an organism that lives in association with a host with a mutual benefit and without harm to either member. **Commensal:** a microorganism that benefits from an association with no known effects on the host. **Pathobiont.** a microbe that does not normally elicit an inflammatory response but under particular conditions (environmentally induced) has the potential to cause dysregulated inflammation and lead to disease. Adapted from Round and Mazmanian, 2009 (*16*).

1.3 Gut microbiota composition: Order in chaos?

The initial gut microbiome population studies, the European metagenomics of the human intestinal tract (MetaHIT) and the US human metagenome project (HMP) have unveiled that gut microbiota composition differs strongly between individuals (*8*, *17*). Despite of the large variation observed in the gut microbiome between individuals, in 2011 the MetaHIT consortia using 33 metagenome-fecal samples from donors distributed in three different continents (Asia, Europe and the United States) proposed

that humans can be clustered into three distinct groups or “enterotypes” based on their gut taxonomic composition (*18*). The authors described their observations as “densely populated areas in a multidimensional space of community composition” these clusters were independent of age, gender, cultural background and geography. Notably, each cluster displayed a similar composition pattern within subjects and a high predominance by one of three different bacterial genera: *Bacteroides*, *Prevotella*, or members of the order Clostridiales, specifically species of the genera *Ruminococcus* and *Fecalibacterium* (**Figure** 1.1).

Reproducible patterns in the predominance of *Bacteroides* and *Prevotella* have been observed in the adult human gut. However, studies using large and geographically diverse cohorts have failed in identifying the three enterotypes (*19*, *20*). A recent study by Gorvitovskaia et al., 2016 (*21*) demonstrated that the concept of a gut “enterotype” could be “misleading” since the clusters observed are the projections of the differences in predominance between *Prevotella* and *Bacteroides*, and it does not reflect consistent patterns of the microbial communities within the “enterotypes”. Specifically, if the sequencing reads belonging to *Prevotella* and *Bacteroides* are removed from the datasets, the clusters are indistinguishable.

Hence, whether the large diversity of the human microbiome can be resumed in a finite number of ecosystems base on specific principles or it is just a random consequence of stochastic colonization events regulated by host genetics and environmental factors, is a concept of debate and actively discussion (*19*–*22*).

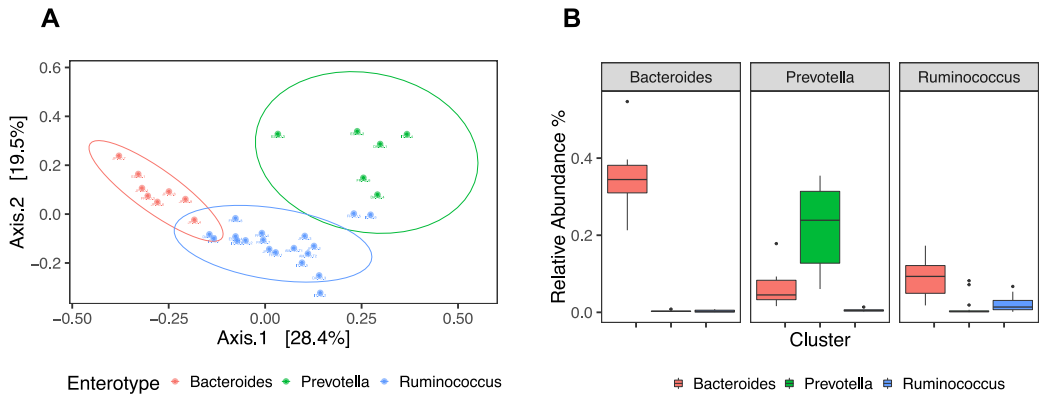

Figure 1.2: Human enterotypes.

Data reproduced from Aruman, et al., 2011. (A) Principal component analysis PCoA. (B) Box-plot of the dominant taxa in each enterotype.

1.4 Major bacterial groups in the gut microbiota

Recent advances in culture-independent techniques, have found a vast diversity in the major gut microbial phyla, Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria, which together represent 98% of the intestinal microbiota. Currently, the most described and cultured species from the gut microbiome belong to three of the main groups of strict anaerobes: *Clostridium cluster XIVa* (also known as the *Clostridium Coccoides* group), *Clostridium cluster IV* (also known as the *Clostridium leptum* group) and *Bacteroidetes* (*23–25*).

1.4.1 Firmicutes: *Clostridia* cluster *IV* and *XIVa*

The class Clostridia is an extremely heterogeneous group of gram-positive spore forming bacteria with several species phylogenetically intermixed. Some Clostridia groups contains significant human pathogenic species, such as *Clostridium tetani*, which are members of *Clostridium cluster I*, and *Clostridioides difficile* belonging to cluster XI. However, most of the intestinal Clostridia have developed a commensal relationship with the host, the clusters *IV* and XIVa comprehend several important commensals for the biology of distinct vertebrates (*26*, *27*).

The taxonomy of *Clostridia* was proposed in the past by Collins et al. 1994 using the 16S rRNA gene (*28*). Since 1994 numerous gut species have been grouped in the clusters *IV* and *XIVa*, but currently this nomenclature does not represent any formal taxonomy such as genus or family. Recently, many of the originally designated species belonging to *Clostridium* clusters have been reclassified into new genera (*29*). However, many species are still misclassified and thereby hold old names witting these clusters. The Clostridium cluster IV includes the genera of *Clostridium*, *Faecalibacterium* and *Ruminococcus*. While the cluster XIVa is composed by species belonging to the genera *Eubacterium, Ruminococcus, Coprococcus, Dorea, Lachnospira, Roseburia and Butyrivibrio.*

Some of the fascinating roles of Clostridia species in the host are the direct and indirect immunomodulatory effects. One of the most interesting host and microbe interactions is the direct effect of Segmented filamentous bacteria or "*Candidatus* Savagella". This bacterium is part of the gut microbiota of rodents, fish and chickens, and has been shown to induce the development of Th17 development, which leads to protection in

specific enteric infection models (*27*, *30*). Additional elegant work by Honda and collaborators describe the indirect effect of 17 strains of Clostridia isolated from the human microbiome on the development of regulatory T cells (Treg) (*26*). The potential mechanism has been associated with the production of SCFA, specifically propionate, acetate and butyrate (*31*, *32*).

1.4.2 Bacteroidetes

The phylum Bacteroidetes is one of the major lineages of Gram-negative bacteria that arose early during the evolutionary process (*33*). This phylum includes non-spore-forming- and rod-shaped bacteria with various abilities to grow anaerobic or aerobic, distributed in the Classes of Rhodothermia, Balneolia, Cytophagia, Sphingobacteria, Chitinophagia, Flavobacteriia and Bacteroidia. The species of Bacteroidetes are widely distributed in the environment and they have been isolated from soil, sediments, marine environments, as well as from the guts and or skin of animals. By far, the members of intestinal *Bacteroidetes* are the most well studied bacteria of the phylum; this group includes the predominant genera of the human gut commensals such as *Bacteroides*, *Prevotella*, *Porphyromonas*, and the less prevalent *Alistipes* and *Parabacteroides*. Notably, the *Bacteroidetes* phylum contains members of the gut microbiota with probiotic, commensal and even pathobiont lifestyle (*34*).

Some of the most evident beneficial role of the Bacteroidetes species is the breakdown of a wide variety of complex polysaccharides. For example *B. thetaiotaomicron*, the first *Bacteroides* species being sequenced (*35*), uses up to one fifth of its genome for starch utilization systems or *sus* genes. The Sus-like systems in combination with distinct

CAZymes form diverse gene complexes called polysaccharide utilization loci (PULs) (*10*, *36*). These molecular systems are highly specific in the recognition of substrates and therefore determine the metabolic niche that *Bacteroides* can occupy (*37*).

Another interesting property of Bacteroides is the anti-inflammatory effect observed in mice colonized with *B. fragilis*. This bacterium expresses a particular zwitterionic capsular polysaccharide (ZPS), which induces T regulatory cells (Tregs) and IL-10-secretion (*38*). Although Bacteroides are mostly associated with promoting health, they could also be associated with risk of disease. One of the interesting ecological mechanisms of *B. thetaiotaomicron* is the enhancement of enteric infections by cleaving host-mucosal glycoproteins into simple sugars, which induces the pathogens virulence factors (*39*).

1.4.3 *Prevotella* (fam. *Prevotella*ceae)

Prevotella is a genus of Gram-negative bacteria of the Bacteroidetes phylum widely spread in distinct body habitats. *Prevotella* strains are non-motile, rod-shaped cells that thrive in anaerobic growth conditions and are traditionally considered commensal since they are rarely involved in infections (*40*). The majority of the known cultured *Prevotella* species have been isolated from the mouth, and a lower number from the gut and the urogenital tract. In the human gut, the species of *P. copri, P. stercorea* and *P. histicola* represent until now the cultivable fraction, with *P. copri* being one of the most common species found (*41*). Several studies have suggested that members of the *Prevotella* genus are beneficial commensals as they are associated

with non-western and rural population that consume a plant-rich diet, as well as with an improvement of glucose responses in individuals with a prebiotic high-fiber diet intervention (*42*–*45*). However, *Prevotella* species have also been associated with chronic inflammatory diseases, such as rheumatoid arthritis and intestinal inflammation in mice (*46*, *47*), yet, the basis of these contradictory observations remains unclear.

The interest into the biology of *Prevotella* started to grow after it has been identified as one of the three biomarkers in the human enterotypes-concept and as key commensals for the assimilation of complex plant polysaccharides. In addition, *Prevotella* species are part of the most wanted human microbiome taxa, with 118 potential species distributed in different body habitats (*48*). However, the low cultivability and the lack of appropriate genetic tools to dissect gene functions have limited the study of this interesting commensal, making it difficult to move from association studies into functional and molecular understanding.

Nevertheless, the significant associations observed between the prevalence of *Prevotella* in health and diseases have opened important questions. First, in health: It remains unclear the functional bases of how *Prevotella* improved glucose and insulin responses upon fiber interventions. Second, in risk of disease: What is the role of *Prevotella* in rheumatoid arthritis? Are specific *Prevotella* species the causal agents of inflammatory disease? Or are multiple species from distinct phyla in combination with *Prevotella* the causal agents of an immunological disequilibrium. Answering these questions is a challenging task, since *Prevotella* isolation from larger cohorts in combination with functional studies in murine models is needed to delineate causal relationships.

1.5 Factors that shape the gut microbiome

Studies have indicated that numerous factors could influence the composition of the gut microbiome. Among the main factors, host-genetics and environment have been shown a significant explanation of the gut microbiota variance (Figure 1.3). Recent studies in twins cohorts, observed that the gut microbiome of monozygotic twins is more similar than that of dizygotic twins (*49–51*), Together these findings supported the significant role of host genetics in shaping the microbiome composition. Furthermore, genome-wide associations studies (GWAS) in combination with microbiome studies (mGWAS) have identified a correlation between human genetic variants and specific taxonomic groups. One of the interesting examples is the observation of an enrichment of Proteobacteria in individuals with SNPs on the oligomerization-domain receptor gene (*Nod2*), a gene highly associated with patients with gastrointestinal pathologies (*49*). Recently, a GWAS study determined that genetics explain approximately 10% of the gut microbiome variance in 1812 individuals from northern Germany (*52*). Whilst environmental factors explain from 10 to 20%, yet, it remains unclear which factors can explain the remaining variance.

Of the environmental factors, diet is one of the most important aspects in shaping microbial communities in the gut (*53–55*) followed by medication (*56*), geography and age (*57*, *58*). The effect of diet on the composition of gut microbiome and human health has been intensively reviewed (*59*, *60*), despite the vast diversity in human alimentary habits, it is becoming more clear that dietary glycans directly influence the prevalence of dominant species in the gut (*35*, *61*). Thus, the detailed dissection and understanding of the relative contribution of the principal factors that shape the gut

microbiome will allow clinicians to design efficient personalized therapies for patients where bacterial dysbiosis contributes significantly to the develop of disease.

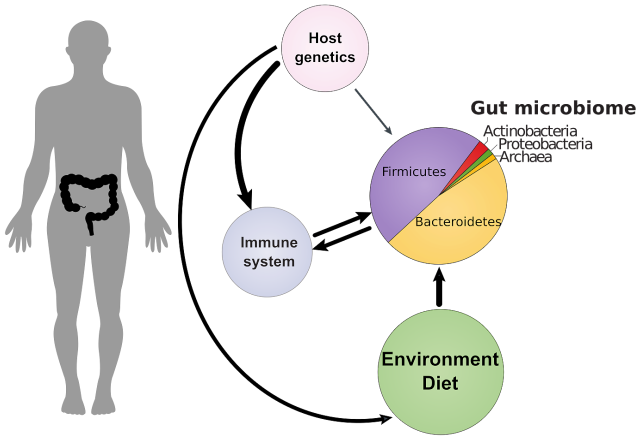

Figure 1.3: Major Factors shaping the gut microbiome
The size of the arrows and shapes indicates the relative contribution.

1.6 Host genetics, microbiome and IBD associations

Inflammatory bowel disease (IBD) are chronic intestinal disorders that include two main types of disease: i) ulcerative colitis (UC) and Crohn's disease (CD). In UC exacerbated inflammation is restricted to the colon, while in CD inflammation can occur between the mouth and the anus compromising inflamed sections mixed in between healthy areas of the intestine. Currently, IBD affects millions of individuals worldwide and its incidence is increasing particularly in the western and westernized populations including pediatric patients (62). Numerous studies have now resulted in a model that IBD is the outcome from an unbalanced cross-talk between the host-genetics, the intestinal microbiota and environmental factors. However, despite several efforts in

understanding the molecular bases and searching for alternative treatments, IBD is still incurable.

In order to gain an advanced understanding of the contribution of host genetics to IBD recent large scale (GWAS) comprising more than 70.000 individuals (35197 healthy, 20155 CD and 15191 UC) have identified significant correlations between human genetic variants and IBD (63). Among the prominent genetic findings, distinct components of the innate immune system (CARD9, NOD2), the interleukin-23-Th17 pathway (IL23R) and genes in associated loci such as IL12B, FUT2 and PTPN22 were significantly enriched. Of the components of the innate immune response, the nucleotide oligomerization domain NOD-like receptors (NLRs) have been particularly strongly associated with Crohn's Disease (64). Especially, single nucleotide variants (SNV) on the NOD2 gene, which encodes an intracellular sensor for recognition of peptidoglycan, a conserved component of bacterial cell walls (65). The association of the NOD2 gene with CD has encouraged further studies to explore the contribution of additional immune sensors of the innate immune system such as the NLR family CARD and PYD domain receptors (NLRCs, NLRPs) and the Toll-like receptors (TLRs). However, these studies in healthy human cohorts have strongly suggested that genomic variants in immune sensors and specifically NLRs are not sufficient to cause CD, pointing to the complexity of IBD pathologies and leading to the investigation of the interplay between host-immune deficiencies, the gut microbiome and diet.

1.7 Aims of this study

The interplay of host and microbiota is driven by complex interactions between microbial, environmental and host-derived factors. Previously, Elinav and Strowig et al., identified and characterized a colitogenic microbiota in mice deficient in the NLRP6 inflammasome. The altered microbiota in these mice was directly responsible to enhance the severity of gut inflammation in a model of experimental colitis and contained a high relative abundance of uncharacterized members of the family *Prevotella*ceae (*47*). Yet, why and how these communities assemble in specific ways remained unsolved and represent a relevant model to investigate the role of host genetics and gut microbiome in the development of intestinal inflammation.

Importantly, association studies have shed light on the connection between gut bacteria and the risk of developing certain inflammatory disease. But, the biggest challenges in gut microbiome research remain the transition from correlation to a functional and molecular understanding. To achieve this aim, it is necessary to implement standardized and manipulatable animal models in genetics and microbiome, followed by the study of the prevalent gut commensals. Therefore, the main focus of my PhD thesis was to investigate the interplay of host factors, specifically the Nlrp6 inflammasome and adaptive immunity, in shaping the gut microbiota using standardized microbiota models. In addition, I aimed to characterize the genomics diversity and functional niche of novel isolates of *Prevotella*ceae, an abundant bacterial family in the intestinal microbiota, which has been associated with health and disease in mice and humans.

To approach these important research questions in my thesis, we developed a comprehensive workflow composed by three steps, i) first, the implementation and standardization of reproducible “omics” and bioinformatic tools, ii) second, the application of these tool to characterize experimental animal models for the study of host and bacteria (i.e., gnotobiotic mice and *Prevotella*ceae isolates) and iii) third, conducting interventions and challenges to our previous characterized model to eventually establish causal relationships (Figure 1.4). Overall, using our study model I aimed to:

a) To evaluate the microbiota composition and stability of enhanced specific pathogen free mice (eSPF) as a suitable model to study host-genetics and gut microbiome interactions.

b) To estimate the relative contribution of host-genetics and environmental factors in the *Nlrp6-/-* mice using the eSPF model and subsequent manipulations.

c) To characterize the genomic diversity and functional niche of novel members of the genus *Prevotella* isolated from mouse gut microbiota including *Nlrp6-/-* mice.

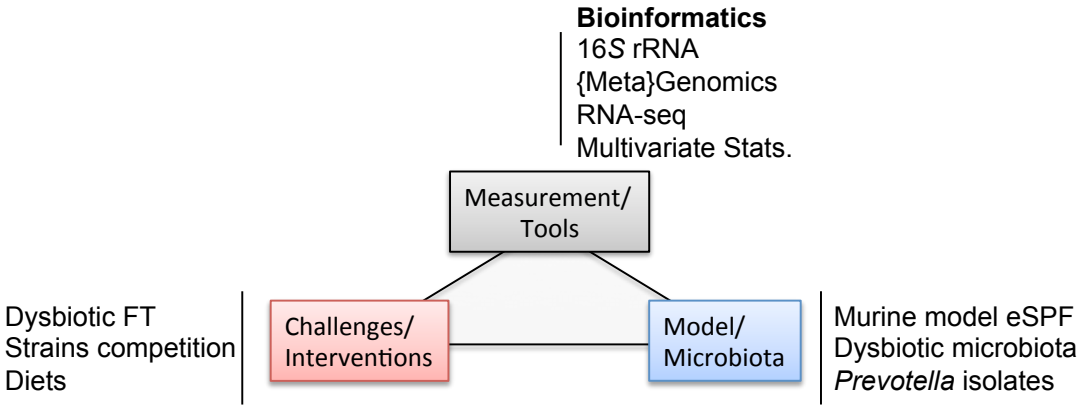

Figure 1.4 Research model

The results of this research are being presented in two subsequent sections. Results section 2.1 describes the effect of host-genetics in shaping the gut microbial composition and section 2.2 presents how environmental factors impact the abundance and functionality of *Prevotella*ceae, a prevalent and uncharacterized gut commensal. The results presented here may help to understand the risk of disease by host genetics and the microbiome, which can be further integrated with broader knowledge for the design of effective diagnostics and treatments directed to target the unbalanced components in the host-microbiome cross-talk.

2 Materials and Methods

2.1 Experimental mouse models

2.1.1 Mouse genetics, housing conditions and rederivation

Colonies of conventional housed WT, *Nlrp6*$^{-/-}$, *Asc*$^{-/-}$, *Casp1/11*$^{-/-}$ and *Rag2*$^{-/-}$ mice on a C57Bl/6N background were bred and maintained within one large animal facility employing standard housing conditions, i.e. housed in individually ventilated cages (IVC) and provided with sterilized food, water and bedding from the same source. The distinct colonies of each genotype originally derived from a single colony of mice, but were subsequently bred and maintained in separated rooms for time periods ranging from several months to years without exchange between the colonies. For the cross-sectional study, samples were taken from 8-12-week-old mice from at least 3 separate cages per colony within a time span of 14 days.

Conventional housed WT, *Nlrp3*$^{-/-}$, *Nlrp6*$^{-/-}$, *Nlrp3*$^{-/-}$*Nlrp6*$^{-/-}$, *Asc*$^{-/-}$ *Casp1*$^{-/-}$, *Casp11*$^{-/-}$, *Casp1/11*$^{-/-}$, *Il1b*$^{-/-}$, *Il1a*$^{-/-}$, *muMT*$^{-/-}$, *Tcrb*$^{-/-}$*Tcrd*$^{-/-}$, and *Rag2*$^{-/-}$ mice on a C57Bl/6N background were rederived using CD1d foster mothers by embryo transfer (ET) at the animal facilities of the Helmholtz Centre for Infection Research (HZI) (Stehr et al., 2009). Foster mothers used for rederivation were taken at different time-points from a continuing and self-contained breeding colony. Foster mothers and offspring obtained after rederivations were maintained and bred under enhanced specific pathogen-free (eSPF) conditions, i.e. housed in IVCs and provided with sterilized food, water and bedding from the same source. Access to animal rooms was restricted to trained animal

caretakers and all manipulations were performed in class II biosafety cabinets. All mice were provided with sterilized food and water ad libitum. Mice were kept under strict 12h light cycle (lights on at 7:00 am and off at 7:00 pm) and housed ingroups of up to 6 mice per cage.

The generation of *Nlrp3*^{tm1Flv} *(Nlrp3*$^{-/-}$*), Nlrp6*^{tm1Flv} *(Nlrp6*$^{-/-}$*), B6.129-Nlrp3*^{tm1Flv}*-Nlrp6*^{tm1Flv} *(Nlrp3*$^{-/-}$*Nlrp6*$^{-/-}$*), Pycard*^{tm1Flv} *(Asc*$^{-/-}$*) Casp1*^{tm2Flv} *(Casp1*$^{-/-}$*), Casp4*$^{tm1a(KOMP)Wtsi}$ *(Casp11*$^{-/-}$*), Casp1*^{tm1Flv} *(Casp1/11*$^{-/-}$*), Il1b*^{tm1Yiw} *(Il1b*$^{-/-}$*), Il1a*^{tm1Yiw} *(Il1a*$^{-/-}$*), Ighm*^{tm1Cgn} *(muMT*$^{-/-}$*), Tcrb*^{tm1Mom}*Tcrd*^{tm1Mom} *(Tcrb*$^{-/-}$*Tcrd*$^{-/-}$*),* and *Rag2*^{tm1Fwa} *(Rag2*$^{-/-}$*)* mice have been previously described.

2.1.2 Mouse vendors and diet interventions

Wild-type (WT) mice in this study were on the C57BL/6N background, colonies were all bred and maintained at the animal facilities of the Helmholtz Centre for Infection Research (HZI) under enhanced specific pathogen-free conditions (eSPF) (*63*). Mice from different commercial vendors were purchased from Janvier, Taconic, Charles Rivers and Harlan (Table 2.2). Germ-free mice were under the wild type C57BL/6NTac background and were maintained in cage isolators (Getinge) in the germ-free facility at the HZI. All experiments were performed with 10- to 14-week-old age-matched and gender-matched animals. Animals were fed with a sterilized standard chow based on plant polysaccharides (Stand-PP). Diet interventions were performed with semisynthetic diet sterilized by gamma irradiation. The used diet in this study were Synth-HF, a high-fat diet (45kJ% fat, lard, SSNIFF D12451) and a control diet Synth-HF, low in fat and

high in sugar (10kJ% fat, 33% sucrose, SSNIFF D12450B). Mice were fed ad libitum and maintained under a strict 12 h light.

Table 2.1: Mice providers

Mouse Line	Genotype	Source: Breeder	Barrier	Commercial
eSPF	C57BL/6NCrl	HZI	A= T2	No
Jan-1a	C57BL/6NRj	Janvier	A= A1	Yes
Jan-10c	C57BL/6NCrl	Janvier	B= C10	Yes
NCI1090	C57BL/6NCrl	National Cancer Ins.	A= T1	Yes
ChR-7	C57BL/6NCrl	Charles River	A=	Yes
Chr-9	C57BL/6NCrl	Charles Rivers	B=	Yes
Chr-11	C57BL/6NCrl	Charles Rivers	C=	Yes
Tac-130	C57BL/6NTac	Taconic	A=	Yes
Tac-401	C57BL/6NTac	Taconic	B= EBU401	Yes
Tac-809	C57BL/6NTac	Taconic	C =	Yes
Har_02	C57BL/6NCrl	Harlan	B =	Yes
N6_DysM	C57BL/6NCrl	HZI	T1	No

2.1.3 Fecal transplantation and sample collection

Microbiome challenging by fecal transplantations was performed according to previously described protocol (*64*). In short, Fecal microbiota transplantation (FT) was done using luminal content of conventionally housed *Nlrp6*$^{-/-}$ mice (*47*), which were derived from “Colony 1” and subsequently transferred to and bred within the conventional barrier of the HZI animal facility without losing their dysbiotic microbiota (*65*). Briefly, mice were euthanized and intestinal content from colon, cecum and small intestine was pooled in anaerobic BBL thioglycolate medium and transferred to an anaerobic chamber (70% N2, 20% CO2, 10% H2). After homogenization the luminal content was filtered through a 70 µm cell strainer and centrifuged for 10 min at 500 g, 4 °C. Bacterial pellet was suspended in anaerobic BHI medium and each WT, *Nlrp6*$^{-/-}$ and

Rag2$^{-/-}$ recipient mouse (males, 6-7 weeks old) received an aliquot of the same preparation by oral gavage. Recipient mice were housed in three separate cages per genotype after FT. Four weeks after FT fecal pellets were collected for DNA extraction as well as luminal content and tissue samples from ileum (SI), cecum (Cec), proximal colon (PC), and distal colon (DC) for RNA isolation and 16*S* rRNA gene amplicon sequencing.

2.2 Experimental procedures

2.2.1 DNA and RNA Isolation as well as cDNA synthesis

For DNA based 16*S* rRNA gene sequencing, fecal pellets were collected and stored at -20 ˚C until processing. DNA was isolated using an established protocol (*66*). Briefly, each sample was treated with 500µl of extraction buffer (200 mM Tris, 20mM EDTA, 200mM NaCl, pH 8.0), 200µl of 20% SDS, 500µl of phenol:chloroform:isoamyl alcohol (24:24:1) and 100µl of zirconia/silica beads (0.1 mm diameter). Samples were homogenized twice with a bead beater (BioSpec) for 2 min. After precipitation of DNA, crude DNA extracts were resuspended in TE Buffer with 100µg/ml RNase I and column purified to remove PCR inhibitors.

For RNA based 16S rRNA sequencing, the gastrointestinal tract was sampled at 4 different sites (Ileum: SI, Cecum: Cec, Proximal Colon: PC and Distal Colon: DC). For each site luminal content and mucosal tissue were collected separately and homogenized in TriReagent (MRC) using a FastPrep-24 Instrument (MP Biomedicals). RNA was isolated according to the manufacturer’s instructions and treated with 2U of

DNase I (Ambion) for 25 min at 37°C. One microgram of total RNA was used to generate cDNA (RevertAid Reverse Transcriptase) using random hexamer primers.

2.2.2 Bacterial strains isolation

For conventionally housed donor *Nlrp6*$^{-/-}$, Janvier and NCI1090 mice, colonic content was collected and homogenized in BBL thioglycolate media and further processed in an anaerobic chamber with following gas mixture: 70% nitrogen, 20% carbon dioxide and 10% hydrogen. *Prevotella* spp., were isolated by using the most probable number (MPN) technique (*67*) where homogenized content was diluted in a range in which maximal 30% of wells showed detectable growth. Specifically, 10-fold dilutions (10^{-6} and 10^{-7}) of fecal content homogenate were cultured in a sterile 96-well plate in Brain Hearth Infusion broth (BHI), supplemented with 10% FBS and 0.5 g/l vitamin (BHI-S) on 37^{o}C for 2 days. *Prevotella*-positive wells were identified using 16S specific primers designed from preliminary metagenomic assemblies (Table 2.1). Positive wells were enriched in BHI-S medium containing vancomycin and BHI-S blood agar plates. Colonies were streaked and isolated 3 times on agar plates before a pure culture was obtained. Bacterial stocks were suspended in 25% of glycerol and furthermore cryopreserved at -80°C. For each experiment, fresh *Prevotella* isolates were grown anaerobically (70% N_2, 20% CO_2 and 10% H_2) from a frozen glycerol stock in BHI-S+ medium on 37^{o}C for 2-3 days. All mice were colonized by oral gavage at age of 5-6 weeks with 200 µl of freshly grown *Prevotella* culture.

Table 2.1: Specific primers for *Prevotella* isolates

Strain	Forward	Reverse	Size bp	Annealing
P. intestinalis	Prevo-N6h-181f 5'-CGTCCCTTGACGGCATCCGACA-3'	Prevo-N6h-1032r 5'-CAGCCCCGAAGGGAAGGGGTG-3'	849	60
P. rodentium	Prevo-JAN-603f 5'-TGAAATGTCGGGGCTCAACCTTGACAC-3'	Prevo-JAN-1289r 5'-GCGGCTTTACGGATTGGACGTACG-3'	686	60
P. muris	Prevo-NCI-61f 5'-GGCAGCATGACATGTTTTCGGACGT-3'	Prevo-NCI-642r 5'-CAGTTCGCGCTGCAGGACCG-3'	579	60
Total *Prevotella* (*68*)	g-Prevo-F 5'-CACRGTAAACGATGGATGCC-3'	g-Prevo-R 5'-GGTCGGGTTGCAGACC-3'	513	55
Total 16S	Uni_334F 5'-ACTCCTACGGGAGGCAGCAGT-3'	Uni_514R 5'-ATTACCGCGGCTGCTGGC-3'	300	55-60

* *Prevotella* species were isolated previously by Aida Iljazovic. Experimental procedures are described in detail in Iljazovic, Aida. 2018. Doctoral thesis TU Braunschweig.

2.2.3 Determination of host gene and protein expression in colon tissue

Colons were excised, washed in PBS and divided into proximal and distal colon. Two centimeter of distal colon was cut longitudinally into two samples: one for RNA and other for protein extraction. RNA isolation and cDNA synthesis were performed as described above. Realtime-PCR was done using Kapa Probe Fast qPCR kit (Kapa Biosystems) and gene-specific probe sets (*Il18* Mm_00434225_m1; *Casp1* Mm_00438023_m1; *Nlrp6* Mm_00460229_m1, Applied Biosystems) and *Hprt* (F: CTGGTGAAAAGGACCTCTCG; R: TGAAGTACTCATTATAGTCAAGGGCA; Probe: TGTTGGATACAGGCCAGACTTTGTTGGAT) on a Light Cycler 480 instrument (Roche). PCR conditions were 95^{o}C for 60 s, followed by 40 cycles of 95^{o}C for 3 s and 60^{o}C for 30 s. Data were analyzed using the delta CT method with *Hprt* serving as the reference housekeeping gene.

Protein extraction was performed by mechanical homogenization of distal colon tissue samples in NP-40 buffer, containing protease inhibitors (Complete Mini EDTA-free, Roche), using Mini-Beadbeater-96 (Biospec). Tissue homogenates were further centrifuged (10,000 r.p.m. for 5 min at 4°C) and the supernatants were collected for IL-18 cytokine measurements using IL-18 ELISA kit (MBL) according to the manufacturer´s instruction.

2.3 16S rRNA sequencing for microbiota profiling

Amplification of the V4 region (F515/R806) of the 16S rRNA gene was performed according to previously described protocols (*69*). Briefly, for DNA-based amplicon sequencing 25 ng of DNA were used per PCR reaction (30 µl). The PCR conditions consisted of initial denaturation for 30s at 98°C, followed by 25 cycles (10s denaturation at 98°C, 20s annealing at 55°C, and polymerase extension at 20s at 72°CDNA. Each sample was amplified in triplicates and subsequently pooled. After normalization PCR amplicons were sequenced on an Illumina MiSeq platform (PE250).

2.4 Total DNA isolation and metagenomic library preparation

Total DNA was isolated from stool pellets or bacterial cells using a phenol-chloroform base protocol. Briefly for *Prevotella* isolates, cells were growth in 5ml Brain Hearth Infusion broth (BHI), supplemented with 10% FBS and 0.5 g/L vitamin K (BHI-

S+) on 37ºC (OD 0.4). Then, cells were centrifuged 3,500 rpm and suspended with 500 µl of extraction buffer (200 mM Tris, 20mM EDTA, 200 mM NaCl, pH 8.0), 200 µl of 20% SDS, 500 µl of phenol:chloroform:isoamyl alcohol (24:24:1) and 100 µl of zirconia/silica beads (0.1 mm diameter). Bacterial cell/fecal pellets were homogenized with a bead beater (BioSpec) for 2 min. DNA was precipitated with absolute Isopropanol and finally washed with 70% vol., ethanol. DNA extracts were suspended in TE Buffer with 100 µg/ml RNase I and furthermore column purified. Total DNA was quantified and diluted to 25 ng/µl.

For Illumina library preparation, 60 µl of total DNA were used for sonication shearing (Covaris). Fragmentation was performed as follow: Processing time = 150sec, Fragment size = 200bp, Intensity= 5, duty cycle= 10. Illumina library preparation was perform using the NEBNext Ultra DNA library prep Kit (New England Biolabs Inc.). The library preparation was performed according to the manufacturer's instructions. I use as input a total of 500ng of sheared DNA; the size selection was performed using AMPure XP beads (First bead selection = 55 µl, and second = 25 µl). Adaptor enrichment was performed using seven cycles of PCR using the NEBNext Multiplex oligos for Illumina (Set1 and Set2) (New England Biolabs Inc.).

2.5 Total RNA Isolation and RNA-seq library preparation

High quality RNA was isolated using acid-phenol chloroform-based protocol [1,3]. Briefly for *in vitro* treatment, bacterial cell were grown in Brain Hearth Infusion broth (BHI), supplemented with 10% FBS and 0.5 g/L vitamin K (BHI-S+) on 37°C to (OD_{600} 0.4). For *in vivo* sampling, intestinal content (~100 mg) was immediately preserved

using bacterial RNA-protect (QIAGEN) and cryopreservation (-80°C). Cell disruption was done using a fast prep (MP) in presence of 200 µL of zirconia beads 0.1 µm diameter and lysis buffer (200 mM NaCl, 20 mM EDTA), 220µL of 20% SDS, 600 µL of phenol:choloroform:isoamyl alcohol (pH 4.5, 125:24:1, Ambion). After centrifugation (12,600 rpm x 5 min at 4°C), the supernatant was discarded and the dry pellets were used for RNA purification with an equal volume of acid phenol chloroform. The total RNA was precipitated using 2 volumes of Isopropanol (>3 hours of incubation) and centrifugation for 30 min 12,600 x rpm at 4°C. RNA pellets were washed with 750 uL of cold EtOH 75% and suspended in 90 µl of 1xTE DNase treatment was done with 2 units of TURBO DNase (Ambion) and furthermore the reaction was purified in silica-based columns using RNeasy Kit (Qiagen).

For meta-RNA-seq library preparation, RNA quality was evaluated using bioanalyzer nano-chip (Agilent technologies), samples were selected according to RNA integrity score (RIN>8.0). Due to the gut microbiota samples contain a high amount of host and bacteria rRNA, I use the Ribo-Zero Gold Epidemiology (Illumina, CA. USA), this kit allowed us to enrich the total microbial messenger RNA (mRNA). After rRNA depletion, the mRNA was fragmented to 200 bp by sonication (Covaris) and evaluated again for quality and size. For each sample a total of 100ng of fragmented mRNA was used as an input for cDNA synthesis and Illumina sequencing adaptor ligation. Illumina sequencing libraries were prepared using a directional RNA Library kit (NEBNext Ultra) (New England Biolabs Inc.) following manufactures' protocol.

2.6 Bioinformatic analyses

2.6.1 16S rRNA gene analysis

Sequencing analysis was performed according to previously described computational workflow (*64*). In brief, obtained reads were assembled, quality controlled and clustered using Usearch8.1 software package (http://www.drive5.com/usearch/). Furthermore, reads were merged using -fastq_mergepairs –with fastq_maxdiffs 30 and quality filtering was done with fastq_filter (-fastq_maxee 1), minimum read length 200 bp. The OTU clusters and representative sequences were determined using the UPARSE algorithm (Edgar, 2013), followed by taxonomy assignment using the Silva database v128 (Quast C, 2013) and the RDP Classifier (Wang et al., 2007) with a bootstrap confidence cutoff of 80%. The operational taxonomic unit (OTU) absolute abundance table and mapping file were used for statistical analyses and data visualization in the R statistical programming environment (http://www.rproject.org)

2.6.2 Genome assembly and annotation

Short reads were obtained for each *Prevotella* species on the Illumina HiSeq 2500 platform. The reads were assembled with SPAdes V.3.10.0 with activated Bayes Hammer tool for error correction and Mismatch Corrector module. Short contigs were then filtered by length and coverage (contigs > 500 bp and coverage > 3X) (*70*) . The resulting contigs were selected for scaffolding using SSPACE (*71*) and furthermore gene prediction and annotation was performed through the Rapid Annotation Subsystem Technology (RAST) server (*72*, *73*).

2.6.3 Differential gene expression RNA-seq analysis

Reads were quality filtered using Trimmomatic with as follow parameters (LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:35 HEADCROP:3). After quality control reads were aligned to *Prevotella* assembled genomes using STAR (*74*). Reads count to each gene was evaluated using HTseq (*75*). Normalization and differential expression were quantified using the DEseq2 package (*76*). Differential expressed gene pathways were identified using KEGG (*77*) and eggnogs (*78*).

2.6.4 Phylogenetic inference

Full 16S phylogenetic analysis was inferred by the GGDC web server (*79*) available at http://ggdc.dsmz.de/. A multiple sequence alignment was created with MUSCLE (*80*). Maximum likelihood (ML) and maximum parsimony (MP) trees were inferred from the alignment with RAxML (*81*) with a 1000 bootstrapping replicates.

2.6.5 Prevalence Analysis using IMNGs-based 16S rRNA amplicon survey

In order to analyze the host prevalence and mean abundance of the *Prevotella* spp., in diverse environments, I used the amplicon data contained in IMNGS (*82*). Using the full-length 16S rRNA gene of each *Prevotella spp*., a similarity search was performed against a local, comprehensive 16S rRNA gene database. Amplicons with > 97% sequence similarity were collected.

2.6.7 Polysaccharide utilization loci (PUL) analysis

PUL prediction was performed base on a similar protocol described in PUL-DB (*83*). In brief, I downloaded the hidden Markov models for SusC and SusD protein domains and furthermore I annotated the Carbohydrate-Active Enzymes (CAZymes) using dbCAN2 meta server with default parameters (*84*). Predicted proteins were considered as positive hits when at least two of the three independent methods (HMMER, DIAMOND, Hotpep) converged with the prediction of a given hit. Then, using the genome coordinates of the predicted *SusC/D* and CAZymes hits; I reconstructed the loci units using a simple moving window consisting of upstream/downstream protein identification with a predicted homologue in the CAZy database. If one homologue was found, the sliding window was moved in that direction, and the search repeated, then the search ended when no more homologues from CAZy were found in the next 5000bp. The PULs were drawn using Python3 and the matplot library.

2.6.7 Statistical analysis

Statistical analyses were performed using R version 3.4.3 (2017-11-30), (http://www.rproject.org) and the packages 'phyloseq' (*85*), and 'ggplot2' (*86*). The permutational multivariate analysis of variance test (ADONIS) and Analysis of Similarities (ANOSIM) were computed with 999 permutations. For ADONIS tests, a $R^2 > 0.1$ (effect size 10%) and P-value < 0.05 was considered as significant. For ANOSIM tests, a R > 0.2 and P-value < 0.05 was considered as significant. To determine differentially abundant (DA) bacterial families, I used linear discriminant analysis (LDA) effect size (LEfSe)(*87*) method with Kruskal-Wallis test <0.05 and LDA scores > 3.5. Then, I implemented DESeq2 with Benjamini-Hochberg multiple testing correction (Love

et al., 2014, McMurdie et al., 2014) to identify differentially abundant OTUs in the gut lumen and mucosa of WT versus gene-deficient mice. OTUs were considered significantly DA between genotypes if their adjusted p-value was below 0.05 and if the estimated fold change was > 2 (*76*). For Mann-Whitney U tests, *P* values lower than 0.05 were considered as significant after correction for multiple testing (Benjamini-Hochberg false discovery rate correction).

3 Results

3.1 Enhanced specific pathogen free [eSPF] mice: A stable and defined gut microbiota model to study host-microbial interactions

Animal models, particularly germfree and gene-deficient mice have been fundamental to not only reveal the influence of the microbiota and its metabolic activities on the physiology of the host, but also to study the complex interplay between the microbiota, environmental factors and host genotype (*89*).

Previously, it has been shown that deficiency in the Nlrp6 inflammasome result in distinct alterations of the microbiota that are disease-exacerbating in mouse models of inflammatory bowel disease (IBD), intestinal tumorigenesis, and metabolic dysfunction (*47*, *90*, *91*). Similar observations were made for other immune sensors such as toll like receptor (TLR) 5 and nucleotide-binding oligomerization domain (NOD) 2 as well as adaptive immunity (*92*–*94*). However, contradictory results have been reported raising doubts whether distinct host-immune deficiencies directly cause aberrations in the gut microbiota or whether they rather reflect the influence of environmental factors such as familial transmission (*95*–*97*). Alternatively, it can be hypothesized that conflicting results are caused by differences in the experimental set-up and factors such as the

composition and assembly of the particular evaluated microbiome. Thus, distinct alterations may not be detectable under all conditions tested. Along these lines, recent reports have challenged the proposed role of the Nlrp6 inflammasome in shaping the microbiome claiming that a generalizable impact of this Nlrp6 was not detectable upon rederivation of these mice into specific pathogen free (SPF) conditions(*95*, *98*). Whether the Nlrp6 inflammasome and other components of the immune systems such adaptive immunity are general regulators of microbiome composition or rather influences the structure of the community only under specific conditions is remain unclear.

In this section, I first characterize the microbiota composition and stability of enhanced specific pathogen free (eSPF) mice, a murine model maintained at the high hygienic barrier at the Helmholtz Centre for Infection Research. Furthermore, using the eSPF mice we investigate the contribution of Nlrp6 and adaptive immunity on the composition of the intestinal microbiome using three distinct approaches. Firstly, we performed a large-scale cross-sectional study of conventionally housed *Nlrp6*$^{-/-}$ mice in order to estimate the influence of genotype, familial transmission and cage effects. Second, we evaluated gene-deficient *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice, which were rederived into standardized (eSPF). Finally, we analyzed the dynamic assembly of the microbiota upon controlled introduction of a dysbiotic community into isobiotic wild type (WT) and gene-deficient mice using 16*S* rRNA gene sequencing to specifically survey the structure and composition of metabolically active bacteria. Our results highlight the large effect of familial transmission on community structure that has not previously demonstrated for Nlrp6 inflammasome deficient mice affecting the presence of numerous bacterial families including Prevotellaceae and Helicobacteraceae. Moreover, characterization of mice raised in eSPF conditions revealed that neither Nlrp6 nor

adaptive immune cells contribute to shaping the microbiota composition under the conditions tested. However, upon reintroduction of a dysbiotic community into isobiotic WT, *Nlrp6*$^{-/-}$, and *Rag2*$^{-/-}$ mice, distinct differences in the composition of the metabolically active microbiota became detectable. We therefore suggest that Nlrp6 as well as adaptive immunity specifically contribute to regulation of microbiota composition in the presence of microbes with proinflammatory features. Additionally, we presented a detailed description of a standardized microbiota (eSPF) as stable and reproducible *in vivo* model for study host-gut microbiota interactions.

3.1.1 Enhanced Specific Pathogen Free [eSPF] mice

In a previous study, Stehr M., *et al.,* 2009 described the stability of the Charles River altered Schaedler flora (ASF) during seven generations in a newly re-derived breeding mouse colony at the Helmholtz Centre for Infection Research (Braunschweig). This study was conducted using cultivable techniques and molecular tests based on the PCR of the reference ASF strains. Here, with the aim to characterize the gut microbiome composition through non-cultivable approaches, we performed 16*S* rRNA gene amplicon sequencing of WT mice from different generations in comparison with ASF mice (Hannover Low-complexity) and a commercial vendor Janvier (Conventional housing) (Figure 3.1.1A).

Non-metric multidimensional scaling (NMDS) ordination analyses indicate three clusters, which correlate with the mice barriers (Figure 3.1.1B). As it was expected, the composition of ASF is composed by 8 dominant OTUs belonging to the Order Clostridiales (n=5), Lactobacillales (n=1), Bacteroidales (n=1) and Deferribacterales

(n=1). Interestingly, the composition of the eSPF microbiome showed unexpected high complexity in comparison with the previous observations by Stehr M., and colleagues. The meta-community contains the eighth bacterial members of ASF microbiota and in addition 71 more OTUs were found (Rel. abundance > 0.5%) grouped in six Orders; Clostridiales (n=69, mean 55.72%), Lactobacillales (n=3, mean=23.86%), Bacteroidales (n=3, mean 16.40%) Deferribacterales (n=1, mean=0.97%), Coriobacteriales (n=1, mean= 1.24%) and Erysipelotrichales (n=2, mean= 0,5%) (Figure 3.1.1C). The highest abundance and diversity was observed in the order Clostridiales, specifically in the genus *Ruminococcus* and unknown OTUs of Lachnospiraceae family. The second most abundant Order observed was Bacteroidales, but interestingly it is composed by just two bacterial genera; *Rikenella* and *Alistipes* (fam. Porphyromonadaceae). Analysis of the low abundance fraction (Rel. abundance <2%), revealed the presence of *Turicibacter* and *Coriobacterium*. In the case of *E. faecalis* and *E. coli* which have been reported as members of ASF microbiota, we found *E. faecalis* under the detection limit (*E. faecalis* < 0.5% rel. abundance) and we did not detect *E. coli* in the screened cohorts. Comparison with samples from a commercial vendor (Janvier) highlighted the absence of common commensal families in the mice gut such as *Prevotella*ceae, Bacteroidaceae and the poorly characterized S24-7 family. Alpha diversity analysis using Shannon index showed that eSPF mice harbors an intermediate diversity (mean = 2.91), which is lower than the conventional Janvier (mean = 3.52), but it is not a low diversity microbiota model as ASF (mean = 1.22) (Figure 3.1.1D). These results strongly support that the eSPF microbiota is devoid of pathogens and it is a valuable model to investigate the effect of potential pathobionts, essentially members of the phylum

Proteobacteria and the little characterized families of Bacteroidetes like *Prevotella* or S24-7 bacteria.

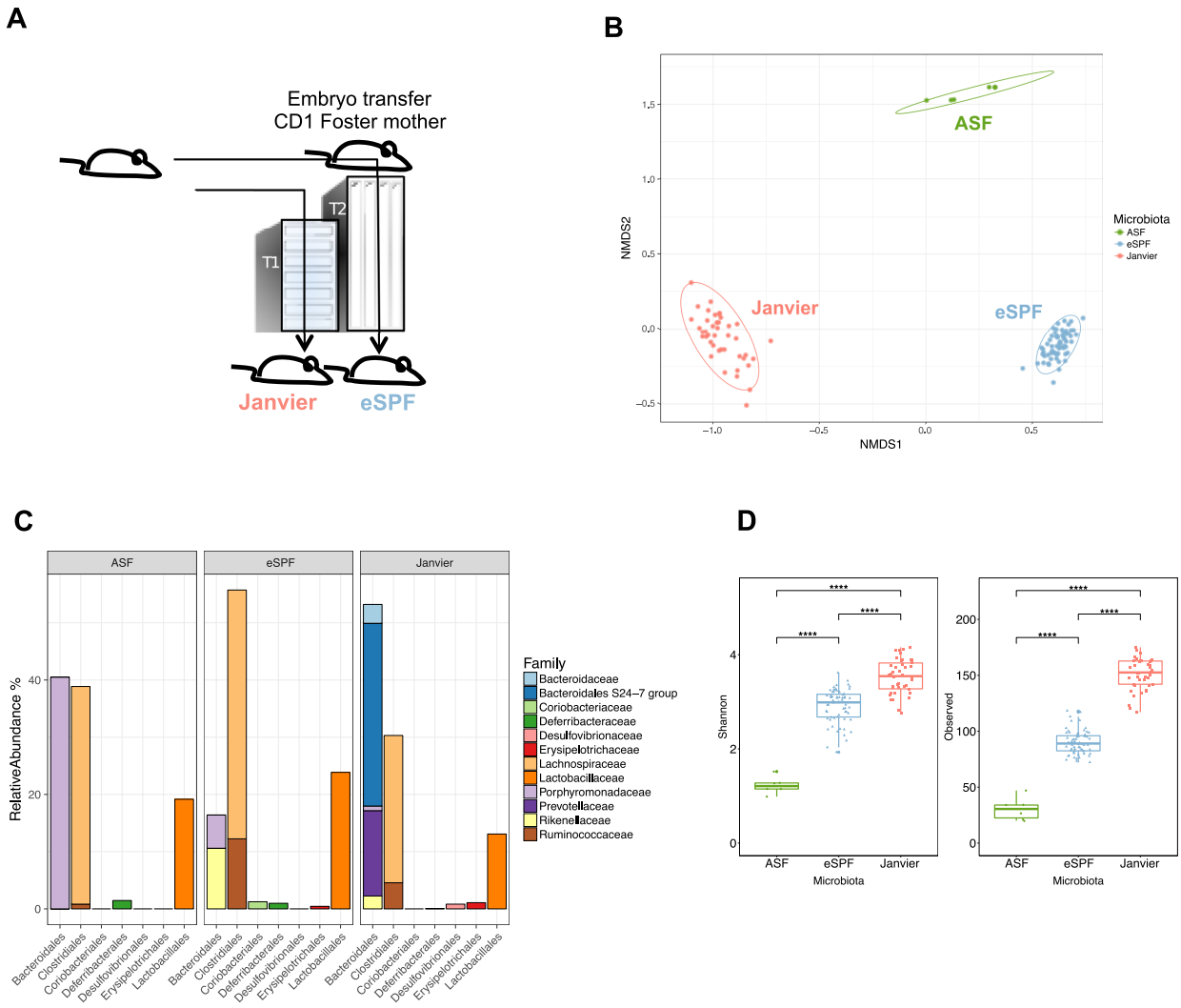

Figure 3.1.1: Gut microbiota composition in enhanced specific pathogen free (eSPF) mice

(A) Diagram of the animal facility barriers at the Helmholtz Centre for Infection Research (Braunschweig), T1 represents the experimental barrier and T2 represents the highly hygienic barrier where the eSPF mice are kept. (B) NMDS ordination analysis of the gut microbiome composition of eSPF vs. low complexity microbiome (ASF) and a commercial vendor (Janvier). (C) Relative abundance of the microbial composition in ASF, eSPF and Janvier. (D) Alpha diversity measurement using observed richness and Shannon index (total n = 111 individuals).

3.1.2 eSPF microbiota composition is independent of deficiencies in adaptive immune cells and Inflammasome components

Thereafter, I aimed to characterize whether deficiencies in inflammasome components (*Nlrp3*$^{-/-}$, *Nlrp6*$^{-/-}$, *Asc*$^{-/-}$, *Casp1*$^{-/-}$, *Casp11*$^{-/-}$, *Casp1/11*$^{-/-}$, *Il1a*$^{-/-}$, *Il1b*$^{-/-}$ mice) or specific subsets in adaptive immune cells (*Tcrb*$^{-/-}$*Tcrd*$^{-/-}$, *muMT*$^{-/-}$, *Rag2*$^{-/-}$ mice) result in genotype-specific clustering in the eSPF microbiota. For this aim, I analyzed the microbiota composition of 336 mice from 13 different lines in the T2 facility rederived in eSPF in comparison with mice housed in a conventional animal house. I observed that samples from individual eSPF mice were interspersed between genotypes without obvious clustering while conventional samples presented a cluster by genotype (Figure 3.1.2A-D). Specifically, genotype did not contribute significantly to the variation observed between WT mice and mice with inflammasome-related deficiencies ($R = 0.040$, $p = 0.106$, ANOSIM-test) as well as deficiencies in adaptive immune cell subsets ($R = 0.154$, $p = 0.002$, ANOSIM) (Figure 3.1.2-A,B). The relative abundance of bacterial families remained comparable between all screened eSPF mice across one year of breeding (Figure 3.1.2E,F). This indicates that under eSPF conditions Nlrp6-dependent effects on the microbiome are not detectable, which is in accordance with the recent study from Mamantopoulos and colleagues (*95*, *98*). However, this observation is not limited to Nlrp6, as it also occurs in mice deficient in adaptive immunity, which have been repeatedly shown to feature microbiome aberrations under conventional housing conditions (*94*, *99*, *100*). Altogether this highlights that eSPF mice represent a stable and reproducible *in vivo* model for study gut microbiota interactions allowing researchers to investigate the influence of key bacterial members on host physiology as well as specific interactions between additional bacteria and the host genotype.

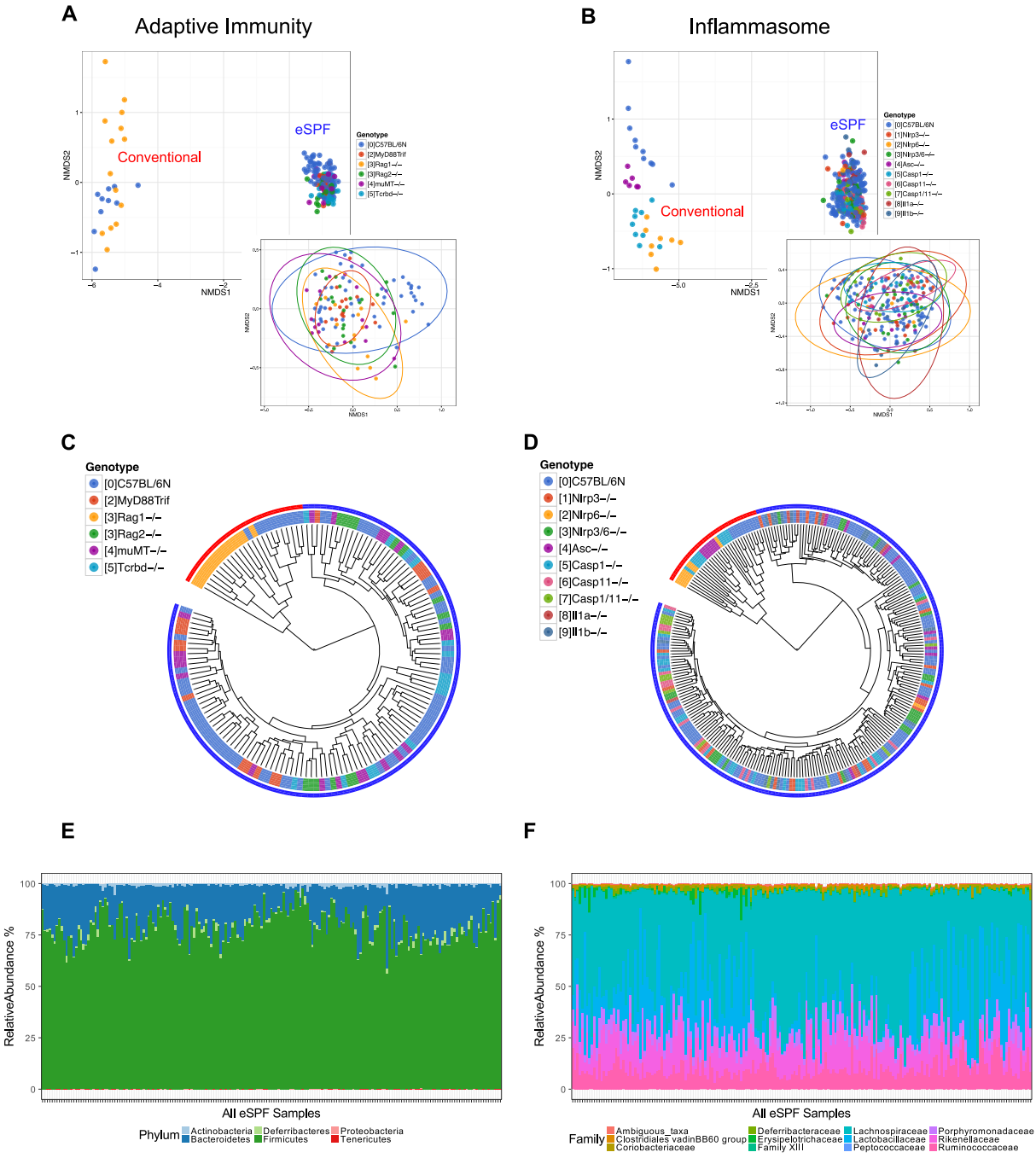

Figure 3.1.2: eSPF microbiota composition is independent of deficiencies in adaptive immune cells and Inflammasome components.

(A and B) NMDS of gut microbiota composition using Bray-Curtis distances in mice deficient of adaptive and inflammasome immune components. (A) NMDS of gut microbiota composition in adaptive and innate knockout mice with standardized eSPF microbiota, stress 0.0593501, ANOSIM p= 0.069. B. NMDS of gut microbiota composition in Inflammasome knockout mice with standardized

(*Legend continued on next page*)

eSPF microbiota, stress 0.06046937, ANOSIM p=0.049. In the ordination analysis the gene deficient mouse strains overlap with WT B/6N (in blue) and genotype-dependent clusters were not observed. For inflammasome components the genotype explained 6% of samples variability (P value = n.s, ANOSIM). (B) Deficiencies in adaptive immunity explained 13% of samples variability (P value = 0.001, ANOSIM). (C and D) UPGMA representation of individual mice samples using weighted UniFrac distances. Outer rings represent microbiota (Red = HCM, Blue= eSPF), Inner ring colors show mice genotype. Left: adaptive and innate deficiencies. Right: Inflammasome deficiencies. (E and F) Gut bacterial composition stability of eSPF microbiota at (E) phylum and (F) family level. The data presented in this section are representative of 336 mice stool samples collected at different time points over a year. Adapted from [Galvez, E.J.C., *et al*., 2017]

3.1.3 Familial transmission significantly influences microbiota composition in conventionally housed Nlrp6 inflammasome deficient mice

Using a cross-sectional study of distinct lines of WT and gene-deficient mice, Flavell and collaborators previously identified an altered microbial community in Nlrp6 inflammasome deficient mice (*47*). Taking into account recent advances in the understanding of the contribution of environmental and experimental variables on microbiome composition (*97*, *101*, *102*), I aimed to estimate the influence of familiar transmission and host genetics in Nlrp6 inflammasome deficient mice. To this end, we surveyed microbiome composition in mice from 13 colonies of Nlrp6-, Asc-, and Caspase-1-deficient mouse lines as well as one colony of WT mice (n=4-10 mice/group) (Figure 3.1.3A).

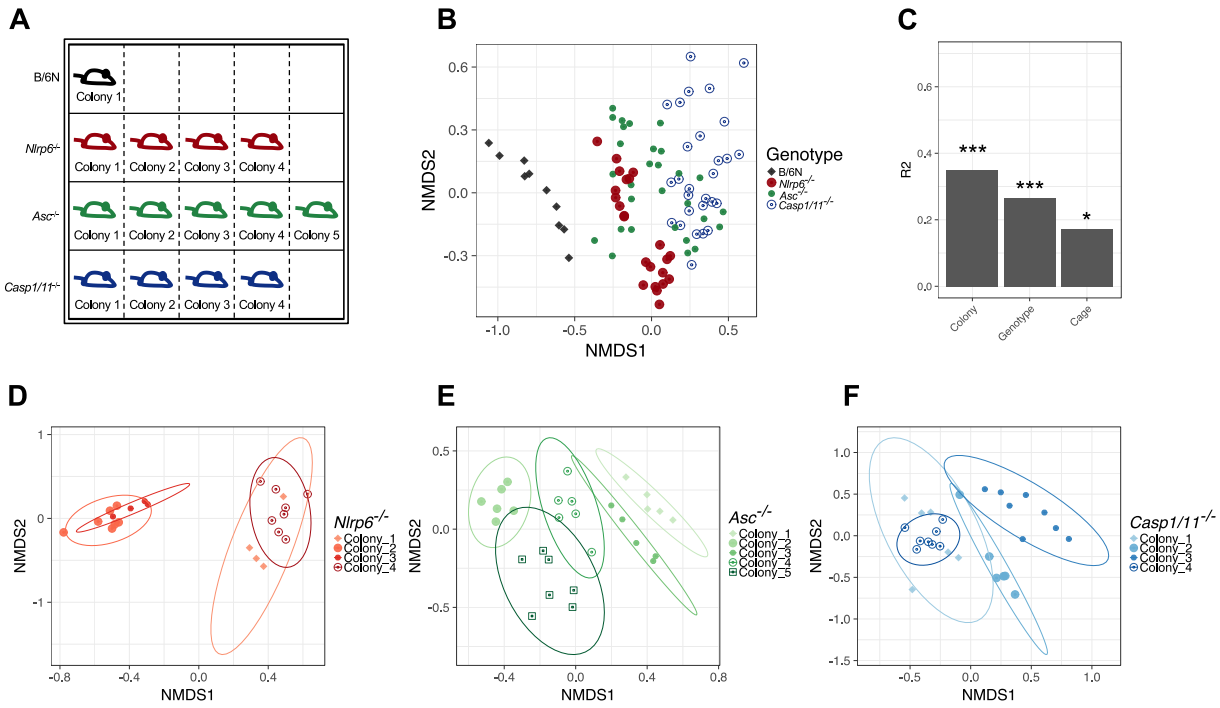

Figure 3.1.3: Environmental factors contribute to diverse microbiota composition in Nlrp6 inflammasome deficient mice.

(A) Scheme for cross-sectional characterization of fecal microbiota composition in WT, *Nlrp6*$^{-/-}$, *Asc*$^{-/-}$, and *Casp1/11*$^{-/-}$ mice maintained by inbreeding in conventional housing conditions. (B) NMDS ordination analysis of microbiota composition in mice using Bray-Curtis distances grouped by genotype (total n = 90 mouse stool samples). (C) Individual effect size of tested covariates. (D-F) NMDS ordination analysis of microbiota composition in *Nlrp6*$^{-/-}$ (D), *Asc*$^{-/-}$ (E), and *Casp1/11*$^{-/-}$ (F) mice using Bray-Curtis distances grouped by colony. Ellipses indicate dispersion of samples within colony. Permutational multivariate analysis of variance (ADONIS) was used to calculate the variance explained by individual factors in C. Adapted from [Galvez, E.J.C., *et al*., 2017]

All the colonies were housed in several different rooms within one large animal facility employing conventional housing conditions (see *Experimental Procedures*). The fecal microbiota was characterized by 16*S* rRNA gene sequencing followed by ordination analyses using Bray-Curtis dissimilarity distances. NMDS revealed a complex pattern of community structures (Figure 3.1.3B). We used permutational multivariate analysis of variance (ADONIS), considering the factors “genotype”, “colony”, and “cage” to estimate their relative contribution to variability within the microbiota. This analysis identified that

gene-deficient mice clustered distinctively compared to WT mice bred and maintained in the same facility (Figure 3.1.3C; R^2 =0.265, p<0.001). But, a greater effect was noted to derive from the factor “colony” (R^2 = 0.349, p<0.001). Notably, the factor “cage” had a lower but significant effect compared to “genotype” (R^2 = 0.172, p<0.05). Finally, when only comparing mice with the same genotype, we confirmed that association with different colonies explained a large fraction of the variability in microbiota composition in *Nlrp6*$^{-/-}$, *Asc*$^{-/-}$, and *Casp1/11*$^{-/-}$ mice (Figures 2.1.3D-F; R^2 = 0.487, 0.555, and 0.408, p < 0.001, respectively).

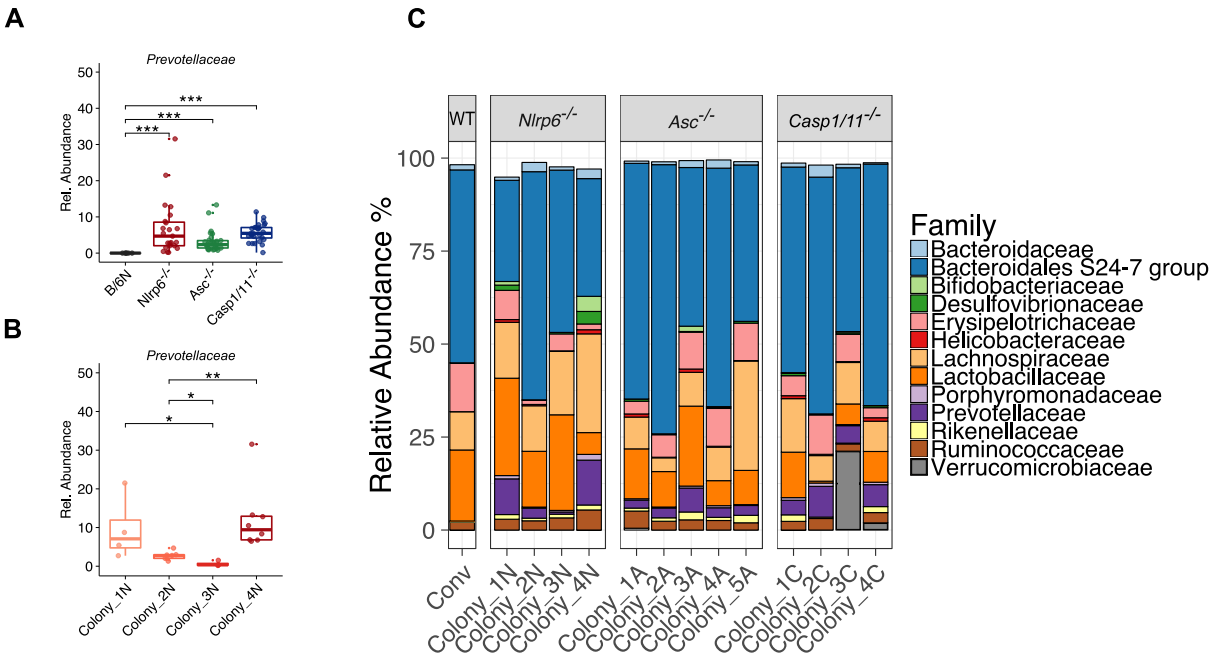

Figure 3.1.4: Environmental factors contribute to diverse microbiota composition in Nlrp6 inflammasome deficient mice.

(A and B) Relative abundance of the family *Prevotella*ceae in mice grouped by genotype (A) and in Nlrp6-/- mice grouped by colony (B). (C) Relative mean abundances of bacterial families in mice grouped by colony. Mann-Whitney U-test was used for A and B with a n = 23-29 mouse stool samples per genotype. ***P<0.001, ***P=0.001 **P=0.01. Adapted from [Galvez, E.J.C., et al., 2017]

The family *Prevotella*ceae, which was observed significantly enriched in *Nlrp6*$^{-/-}$ mice(*47*) had the highest abundance in *Nlrp6*$^{-/-}$ mice compared to WT, *Asc*$^{-/-}$ and *Casp1/11*$^{-/-}$ mice (Figure 3.1.4A), but significant differences were observed in colonies of *Nlrp6*$^{-/-}$ mice (Figure 3.1.4B; $p < 0.001$ Kruskal-Wallis test). Furthermore, OTUs belonging to the genus *Prevotella* differed strongly between mouse lines and even colonies further supporting the high variability in the microbiome of these lines (Figure 3.1.5A). Beside the genus *Prevotella*, similar colony-specific alterations in the abundance of distinct bacterial families were observed in many colonies of gene-deficient mice (Figure 3.1.4C), e.g. high abundances of Verrucomicrobiaceae and Helicobacteraceae in individual colonies, but that were largely absent in other colonies (Figure 3.1.4B-C). Importantly, this cross-sectional study of mice bred under conventional housing conditions highlighted that segregation of mouse lines even within the same facility and kept under the same hygiene conditions has a tremendous impact of microbiota composition such as the presence or absence of a putative pathobionts including *Prevotella spp, Helicobacter spp.* and *Akkermansia muciniphilia* (Figure 3.1.5A). This supports the model that beyond genotype, additional factors such as maternal inheritance and stochastic events influence the composition of the microbiome in conventionally housed mice even when they are maintained within the same facility.

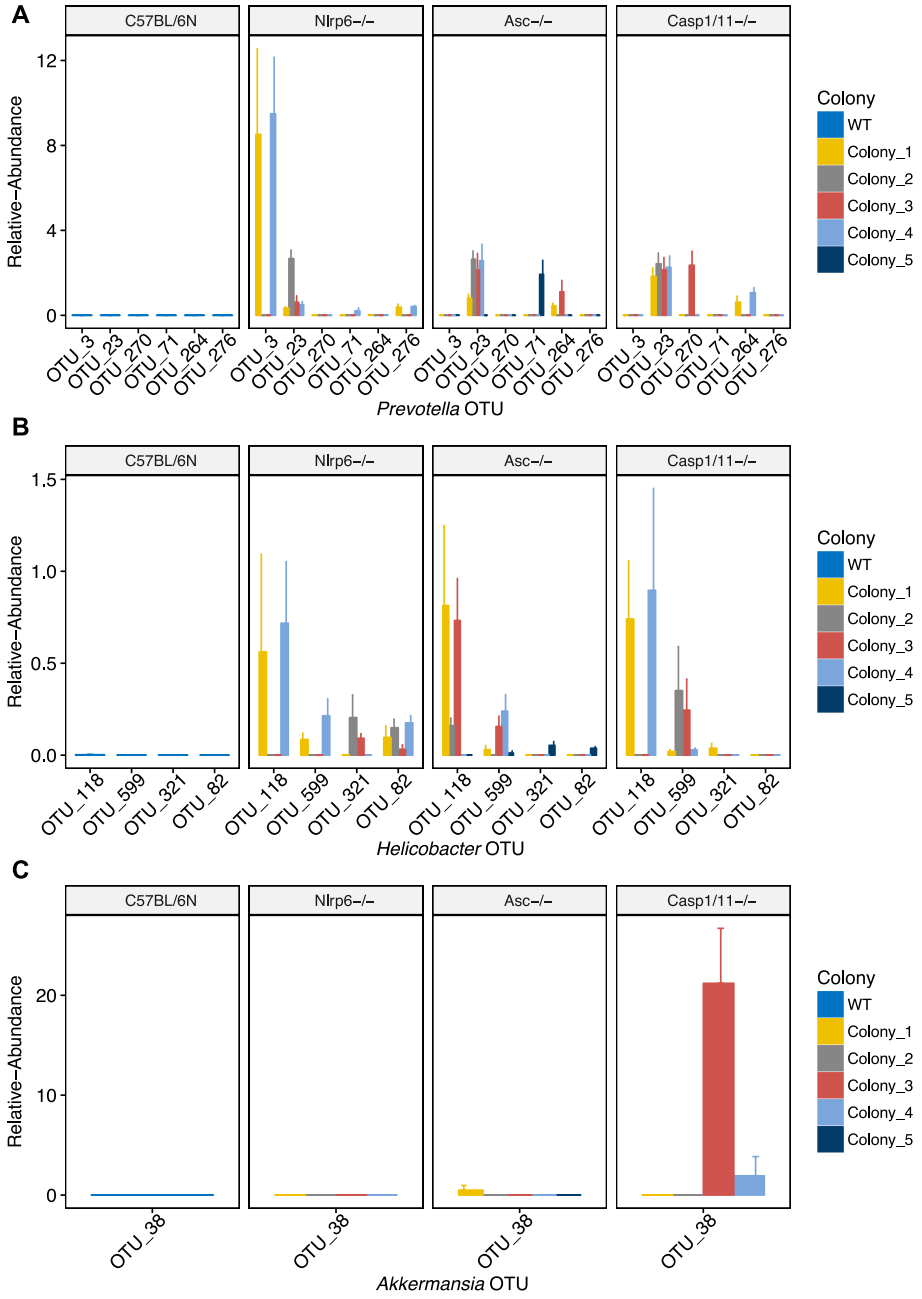

Figure 3.1.5: Conventional housed WT, *Nlrp6*$^{-/-}$, *Asc*$^{-/-}$, and *Casp1/11*$^{-/-}$ mice harbour various relative abundances of distinct OTUs from the families *Prevotella*ceae, Helicobacteraceae and Verrucomicrobiaceae.

(A-C) All OTUs belonging from the families *Prevotella*ceae (A), Helicobacteraceae (B) and Verrucomicrobiaceae (C) were identified and their relative abundance was analyzed in WT (1 colony), *Nlrp6*$^{-/-}$ (4 colonies), *Asc*$^{-/-}$ (5 colonies), and *Casp1/11*$^{-/-}$ (4 colonies) mice.

3.1.4 Rederivation into enhanced SPF conditions diminishes the effect of Nlrp6 deficiency on microbiota composition

To overcome microbiota variability in animal experimentation, it has been proposed standardization of microbiota composition under controlled environmental conditions (*103*). In order to assess the microbiota composition in WT and gene-deficient mice, we employed rederivation of mice by embryo transfer (ET) through a cohort of foster mothers with the eSPF microbiota (Figure 3.1.6A, see methods). Specifically, we focused on two strains of gene-deficient mice, the aforementioned *Nlrp6*$^{-/-}$ mice as well as mice deficient in Rag2 (*Rag2*$^{-/-}$) thereby lacking adaptive immunity. The latter strain was since previous studies had reported conflicting results on the type of alterations in microbiota composition (*94*, *99*). Then, we compared the composition of the fecal microbiota in two cohorts of WT, *Rag2*$^{-/-}$ and *Nlrp6*$^{-/-}$ mice each, one cohort before and one cohort after ET. NMDS analysis showed two main clusters corresponding to the samples before (Conv) and after ET (eSPF) (Figure 3.1.6B). In concordance with previous observations, mice maintained under conventional housing (Conv) clustered by genotype. In contrast, rederivation of *Rag2*$^{-/-}$ and *Nlrp6*$^{-/-}$ mice resulted in offspring with no obvious differences in gut microbiota composition as compared to WT mice (Figure 2.6B). Variability between samples were calculated using Bray-Curtis distances to the respective centroids (eSPF and Conv) demonstrating that beta diversity within eSPF mice is significantly lower than within Conv mice (Figure 3.1.6C; $p<0.001$, Mann–Whitney test). Interestingly, mice with the eSPF microbiota have a similar alpha diversity compared to Conv mice (Figure 3.1.6D). But, while the eSPF microbiota contains mainly Firmicutes and only two families of Bacteroidales (Rikenellaceae and Porphyromonadaceae), Conv mice contain a larger fraction of

distinct Bacteroidales families, e.g. *Prevotellaceae* and S24-7 as well as several representatives from the phylum Proteobacteria, including pathobionts from the families Desulfovibrionaceae, Alcaligenaceae, and Helicobacteraceae (Figure 3.1.6E) (*104*, *105*).

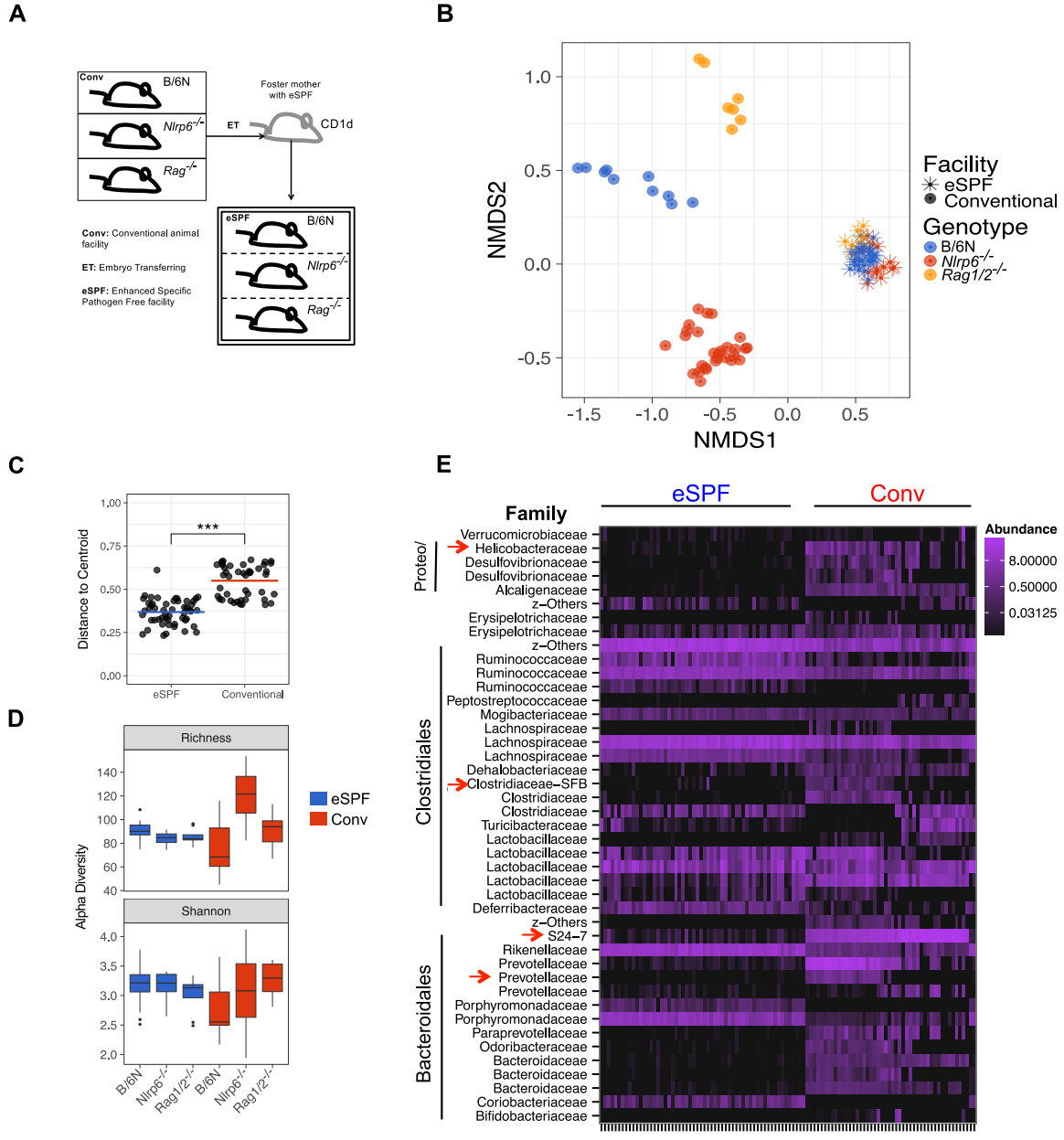

Figure 3.1.6: Rederivation of immunodeficient mice from conventional housing conditions to enhanced SPF conditions normalizes the microbiota.

(A) Scheme for rederivation of conventionally-housed (Conv) mouse strains into enhanced SPF (eSPF) housing conditions via embryo transfer using CD1d foster mothers with eSPF microbiota. (B-E)

(*Legend continued on next page*)

Fecal samples were obtained from Conv- (circles) and eSPF-housed (in asterisks) mice of indicated genotypes. (B) NMDS ordination analysis of microbiota composition using Bray-Curtis distances. (C) Similarity analysis of microbiota composition within housing condition using Bray-Curtis distances to the respective centroid. (D) Estimation of α-diversity using Richness and Inverse Simpson index. (E) Heatmap of detected OTUs clustered at family level, the samples were grouped by housing condition (Microbiota) and genotype as indicated. Mann-Whitney U-test was used in (C), P-value = 0.001, with a total of (n=97) mouse feces samples.

3.1.5 Introduction of potential pathobionts into immunodeficient mice reveals host genetics-driven effects on the microbiome

I hypothesized that unaltered microbiota composition in immunodeficient mice housed in eSPF conditions may be due to the absence of distinct bacteria that are able to explore specific ecological niches opened as a consequence of altered host immune function. Therefore, we introduced the microbiota of conventionally housed dysbiotic *Nlrp6*$^{-/-}$ mice (Conv-Dys) (*47*) into eSPF WT, *Rag2*$^{-/-}$ and *Nlrp6*$^{-/-}$ mice through fecal transplantation (FT) (Figure 3.1.7A). By comparing the fecal microbiota composition, before (eSPF) and after FT (eSPF+Conv−Dys), we observed two main clusters. The first cluster included mice before challenge and the second comprised of mice receiving the FT as well as the donor mice (Figure 3.1.7B). Analysis of OTUs present in recipients and donor mice revealed that 86% of the OTUs from the donor community were being transferred into all three recipient lines (Figure 3.1.7C). Within the cluster of FT recipient mice, further sub-clustering was observed (Figure 3.1.7D), and pair-wise comparisons between WT and gene-deficient mice revealed no significant genotype-driven difference between WT and *Nlrp6*$^{-/-}$ mice using fecal DNA after 4 weeks, but between WT and *Rag2*$^{-/-}$ mice (Figure 3.1.7E, F).

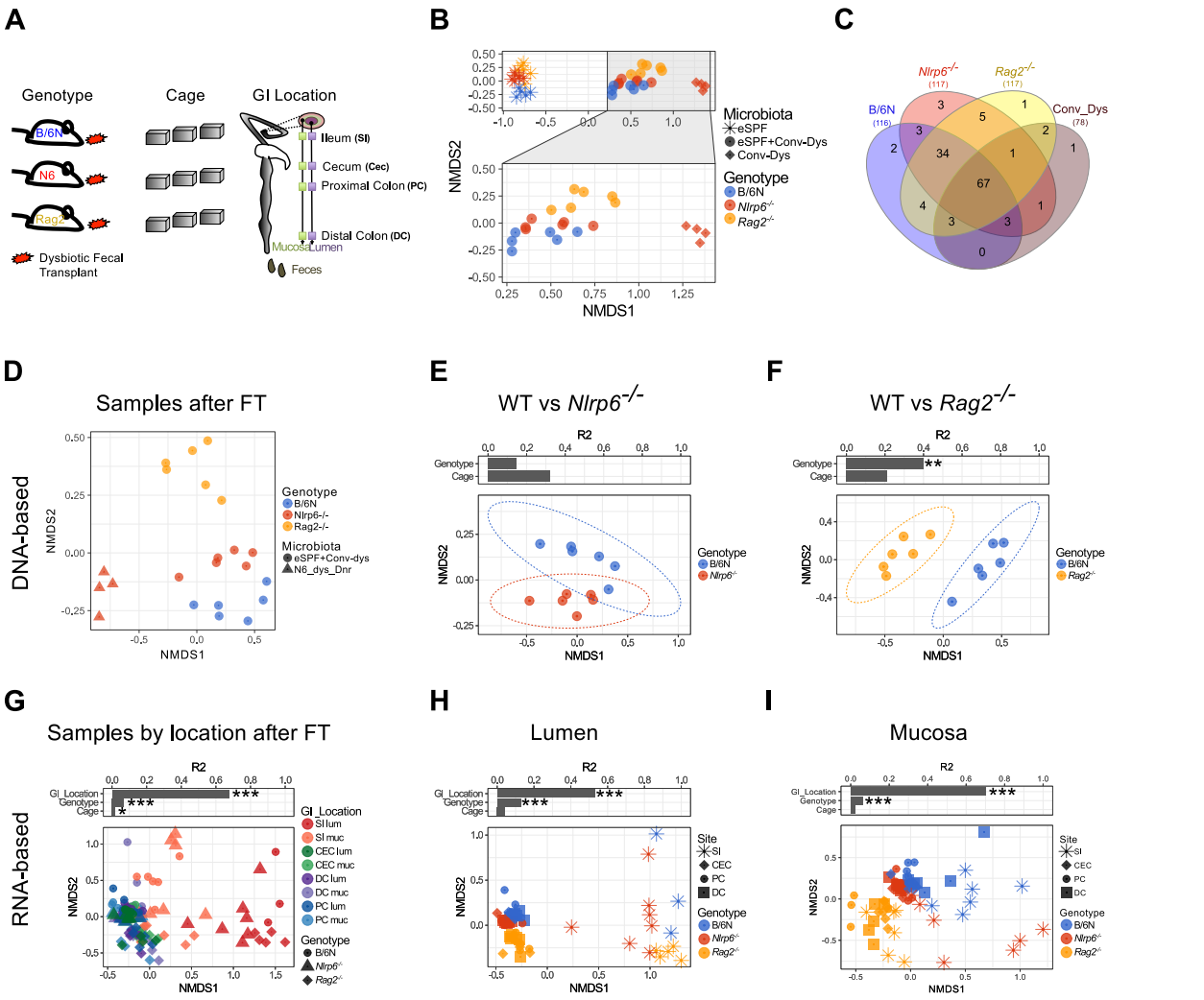

Figure 3.1.7: Gastrointestinal location and host genotype regulate spatial organization of the gut microbiome

(A) Scheme for dysbiotic microbiota transfer experiment. WT (B/6N), *Nlrp6*$^{-/-}$ (N6) and *Rag2*$^{-/-}$ (Rag2) mice bred in eSPF conditions were subjected to a fecal transplant (FT) from conventionally housed *Nlrp6*$^{-/-}$ mice. Fecal samples were taken before and 4 weeks after FT for DNA isolation. Luminal and mucosa-associated samples for RNA isolation were taken from indicated locations (GI locations) 4 weeks after FT. (B) NMDS ordination analysis of microbiota composition using Bray-Curtis distances from fecal DNA of donor *Nlrp6*$^{-/-}$ mice (Conv−Dys) and recipient mice before (eSPF) and after FT (eSPF+Conv−Dys). The zoomed section includes mice after FT as well as donor mice. (C) Venn diagram indicating the number of shared OTUs between the donor mice and the FT recipients. (D, E and F) NMDS of fecal samples after FT using DNA as template. NMDS analysis and individual effect size of “Genotype” and “Cage” using pairwise comparison in *Nlrp6*$^{-/-}$ vs WT (E) and *Rag2*$^{-/-}$ vs WT mice (F). (G, H and I) NMDS of samples across the gastrointestinal tract after FT using RNA as template. NMDS ordination and effect size of GI location, genotype and cage in all samples (G) or from lumen (H) and mucosa (I). Permutational multivariate analysis of variance (ADONIS) was used calculate the variance explained by individual factors in (E) and (F). In total 18 individuals were sampled (n = 6 per genotype). The experimental design and data analyses were performed by Galvez, E.J.C. Iljazovic A, contributed with FT (panel A). Adapted from [Galvez, E.J.C., *et al.*, 2017].

Since Nlrp6 has been proposed to specifically modulate the mucosal barrier (*106*, *107*), I next investigated the microbiota composition in the intestinal lumen or the mucus layer of *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice. Additionally, in order to distinguish metabolically active from inactive bacteria with high sensitivity, we performed 16*S* rRNA gene sequencing based on RNA and DNA (*108–110*). RNA was isolated from luminal and mucosa-associated bacteria at four different anatomic sites, i.e. ileum (SI), cecum (Cec), proximal (PC) and distal colon (DC). Comparative analysis of DNA- vs RNA-based microbiota analysis is described in Figure 2.8.

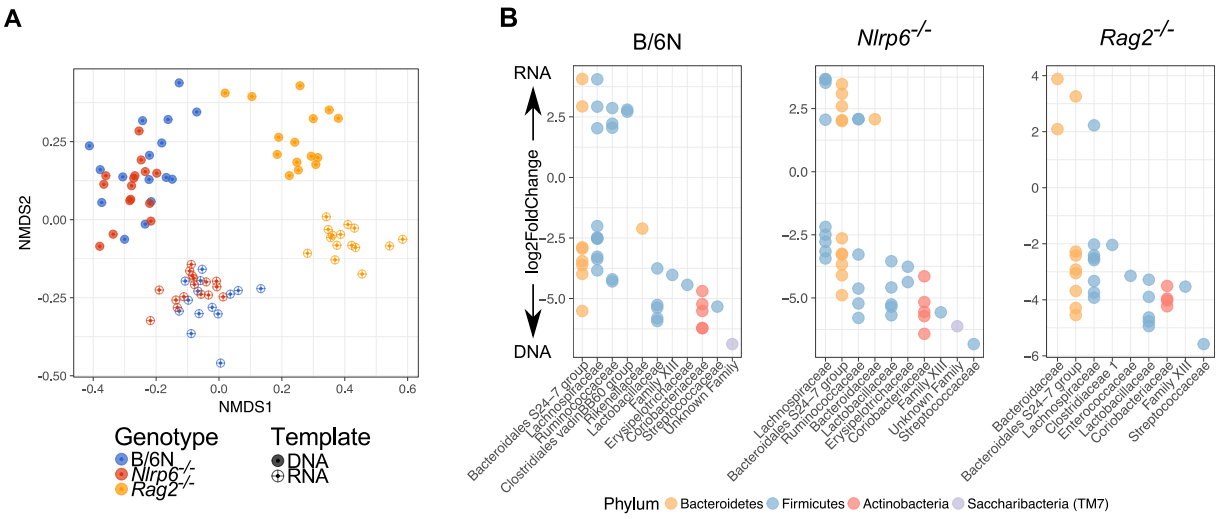

Figure 3.1.8: Comparison of RNA- and DNA-based characterization of the intestinal microbiome.

(A). Left panel: NMDS ordination analysis comparing the microbial composition using RNA (Lumen distal colon) vs DNA (Feces) as template. (B) Right panel: Differential abundant OTUs (Log2fold>4). Adapted from [Galvez, E.J.C., *et al*., 2017]

As expected, analysis of 16*S* rRNA gene sequencing data using NMDS showed several distinct clusters reflecting the different genotypes and anatomical sites (Figure 3.1.7G). Testing with ADONIS indicated that the variable “GI Location” contributes 67% to the

observed variability (R^2 = 0.675, p < 0.001), followed by the variable “Genotype”, which explained 7% (R^2 = 0.070, p < 0.001) (Figure 3.1.7G). Analogous effects were observed when separating samples into luminal and mucosa-associated bacteria with the community in Rag-deficient mice clustering distinctively from WT and *Nlrp6*$^{-/-}$ mice (Figure 3.1.7H and I). After FT, the microbiota was dominated by Clostridiales, Bacteroidales, Campylobacterales and Deferribacterales, which represent 98% of the bacteria found at different GI sites (Figure 3.1.9).

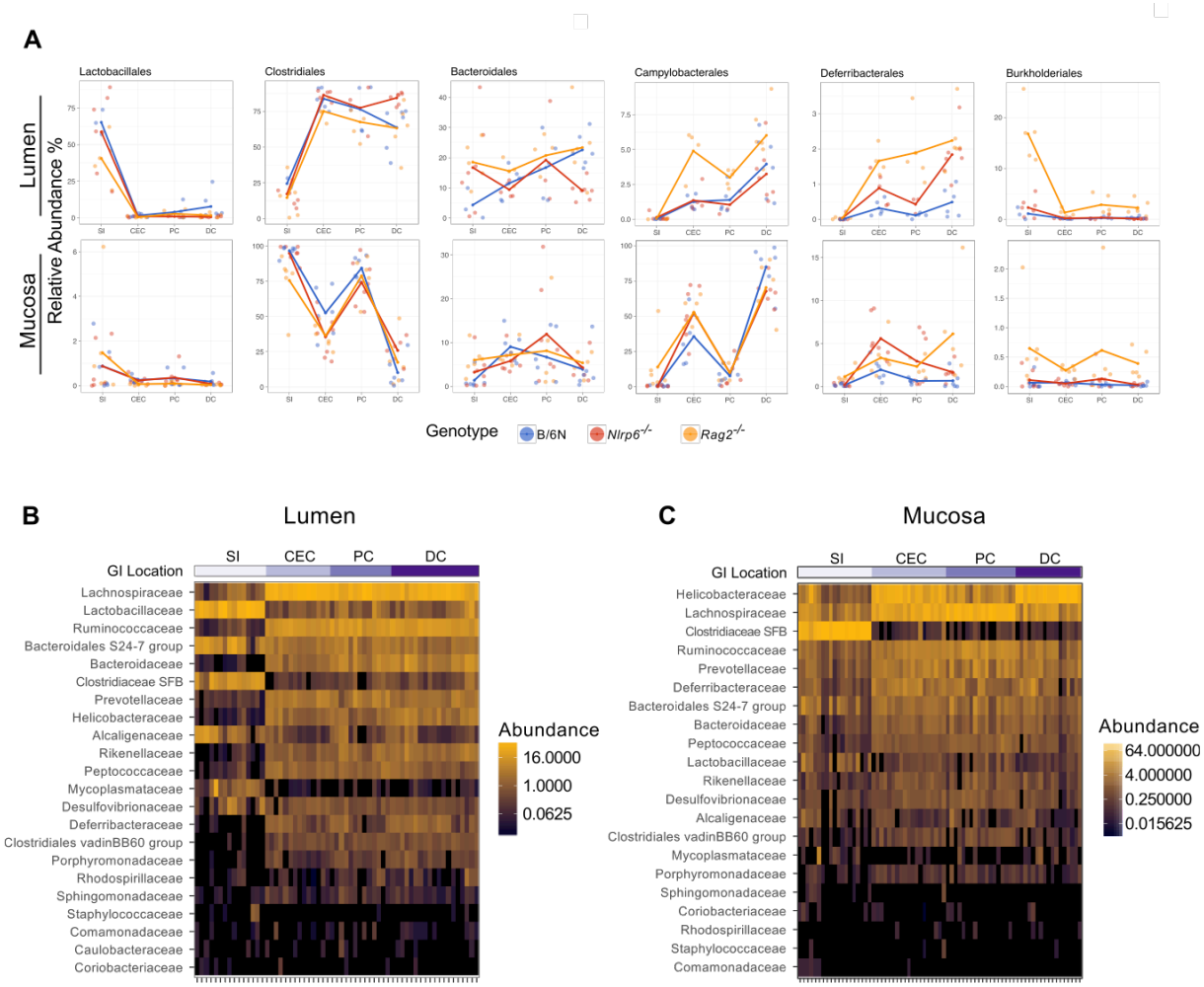

Figure 3.1.9: RNA- and DNA-based intestinal microbiota biogeography 1 in gene-deficient mice.

(A) Analysis of the distribution patterns of the most abundant luminal (upper panels) and mucosa-associated (lower panels) bacterial orders in WT, *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice after transfer of dysbiotic

(*Legend continued on next page*)

community (eSPF+Conv−Dys). X-axis indicates the section in the GI location (SI: small intestine, CEC: cecum, PC: proximal colon, DC: distal colon); Y-axis values indicate the mean relative abundances in each genotype. (B and C) Heat map of bacterial families in lumen (B) and mucosa (C), sorted by relative abundance. Adapted from [Galvez, E.J.C., *et al*., 2017]

In the Lumen, “GI Location” and “Genotype” have a significant effect (R^2= 0.528 and R^2= 0.127, p< 0.001). In the Mucosa, “GI Location” presented a higher contribution in comparison with the lumen (R^2= 0.699, p< 0.001) and “Genotype” also contributed to a lower degree to the differences (R^2= 0.0615, p< 0.001). Pair-wise comparisons of WT and *Nlrp6*$^{-/-}$ mice revealed that communities in the colon but not cecum and SI, differed significantly with “Genotype” explaining up to 41% of variability, i.e. in the lumen of the distal colon (Figure 3.1.10A). Communities in *Rag2*$^{-/-}$ mice differed at most sites except the lumen of the small intestine and the mucosa in the distal colon from the ones in WT mice (Figure 3.1.10B). This demonstrates that upon exposure to pathobiont-containing communities, deficiencies in adaptive immune cells have a broad impact on microbiota composition, while deficiency in Nlrp6 has a rather distinctive, but readily detectable influence on the metabolically active communities in the colon.

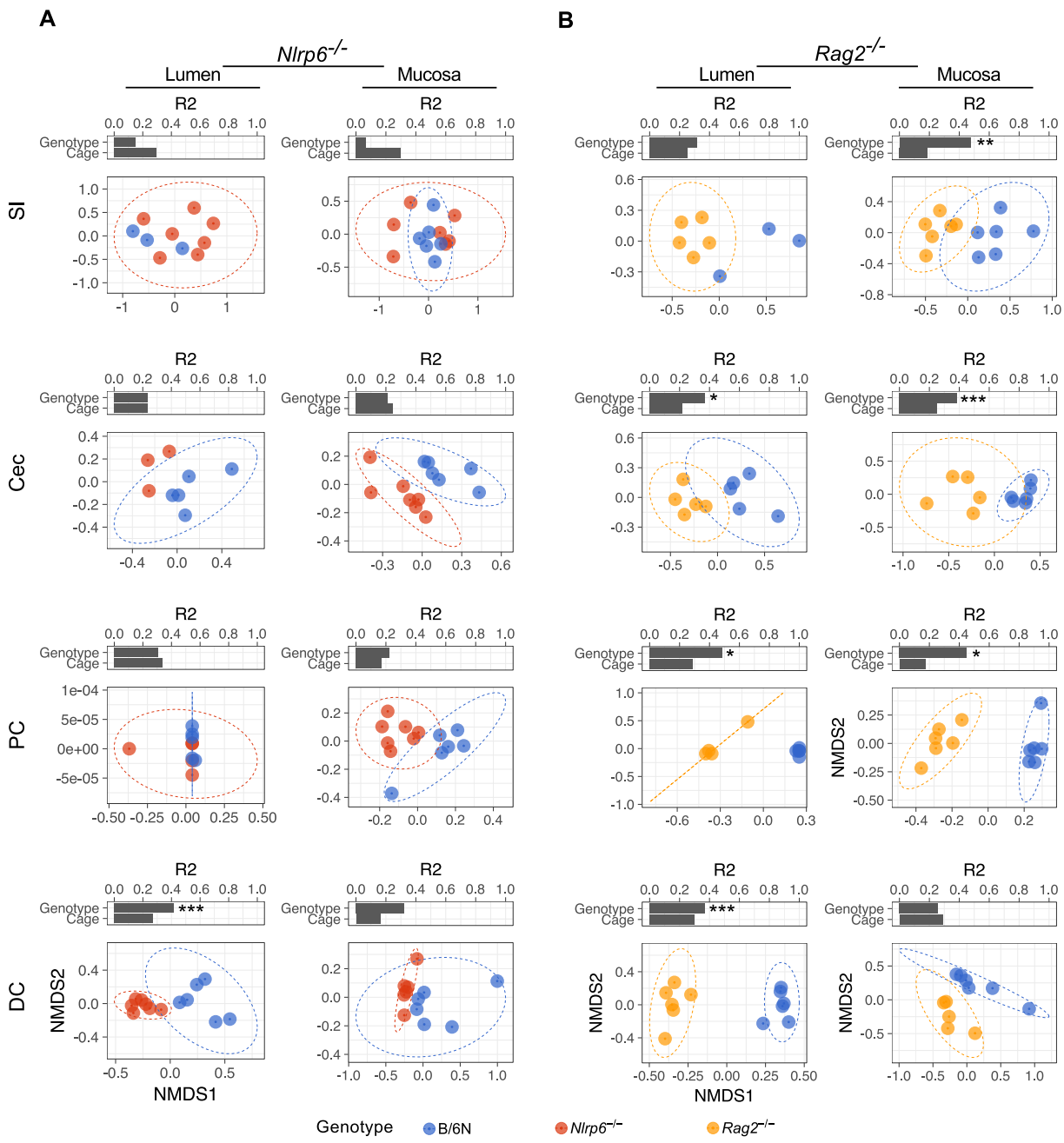

Figure 3.1.10: Location and immune function influence the occurrence of genotype dependent effects on the gut microbiota.

PCoA of gut microbiota composition using Bray-Curtis distances with samples split according to anatomical site. Colors indicate genotype and ellipses indicate the confidence interval, which contains 95% of the samples under a normal distribution. Permutational multivariate analysis of variance (ADONIS) was used to calculate the variance explained by individual factors. A significant effect was attributed when P-value is < 0.05 and R2 is > 0.10 (equivalent to 10% of explained variance); ***P=0.001 **P=0.01, *P=0.05. Adapted from [Galvez, E.J.C., *et al.*, 2017]

3.1.6 Increased abundance of pathobionts in mice deficient of Nlrp6 and adaptive immunity

I next explored the taxonomic composition of the bacteria, which differ between genotypes in the lumen of the colon. For this analysis I included data from two independent FT experiments, i.e. transfer of Conv-DysM microbiota into eSPF mice (WT, *Rag2*$^{-/-}$ and *Nlrp6*$^{-/-}$), and employed complementary statistical approaches to identify biomarkers at higher taxonomic levels by the linear discriminant analysis (as described in LEfSe (*87*) and differentially abundance (DA) OTUs using the negative binomial Wald test (as described in (*76*). Initial NMDS analysis confirmed that samples readily clustered by genotype in the lumen of the proximal and distal colon, which was corroborated by statistical analysis (ADONIS, $R^2 > 0.10$; $P < 0.01$) (Figure 3.1.11A-D). I next performed LEfSe analysis (LDA score > 3.0) to identify DA bacterial families at the different sites in the colon. In the DC of *Nlrp6*$^{-/-}$ mice the families of Helicobacteraceae, Deferribacteraceae and Desulfovibrinaceae were enriched, while Lactobacillaceae and Bacteroidaceae were reduced (Figure 3.1.11B). A similar observation was made for the DC of *Rag2*$^{-/-}$ mice (Figure 3.1.11D). Notably, a different pattern was detected in the PC of *Nlrp6*$^{-/-}$ mice showing an enrichment of Porphyromonadaceae and a decrease in the little described family/cluster Clostridiales–vadinBB660 (Figure 3.1.11A), while the DA bacteria in the PC of Rag mice resembled the ones in the DC (Figure 3.1.11C). I then performed analysis of DA OTUs using DESeq2 (log2FoldChange > 2.0, $P < 0.05$ after correction for multiple test) for the PC and DC comparing each gene-deficient mouse line separately against WT mice. For *Nlrp6*$^{-/-}$ mice we identified 23 (PC, 13 up and 10 down) and 24 (DC 15 up and 9 down) distinct DA OTUs (Figure 3.1.11A and 11B). In

Rag2$^{-/-}$ mice, we identified 86 (PC, 26 up and 60 down) and 70 (DC 22 up and 48 down) distinct DA OTUs (Figure 3.1.11C and D).

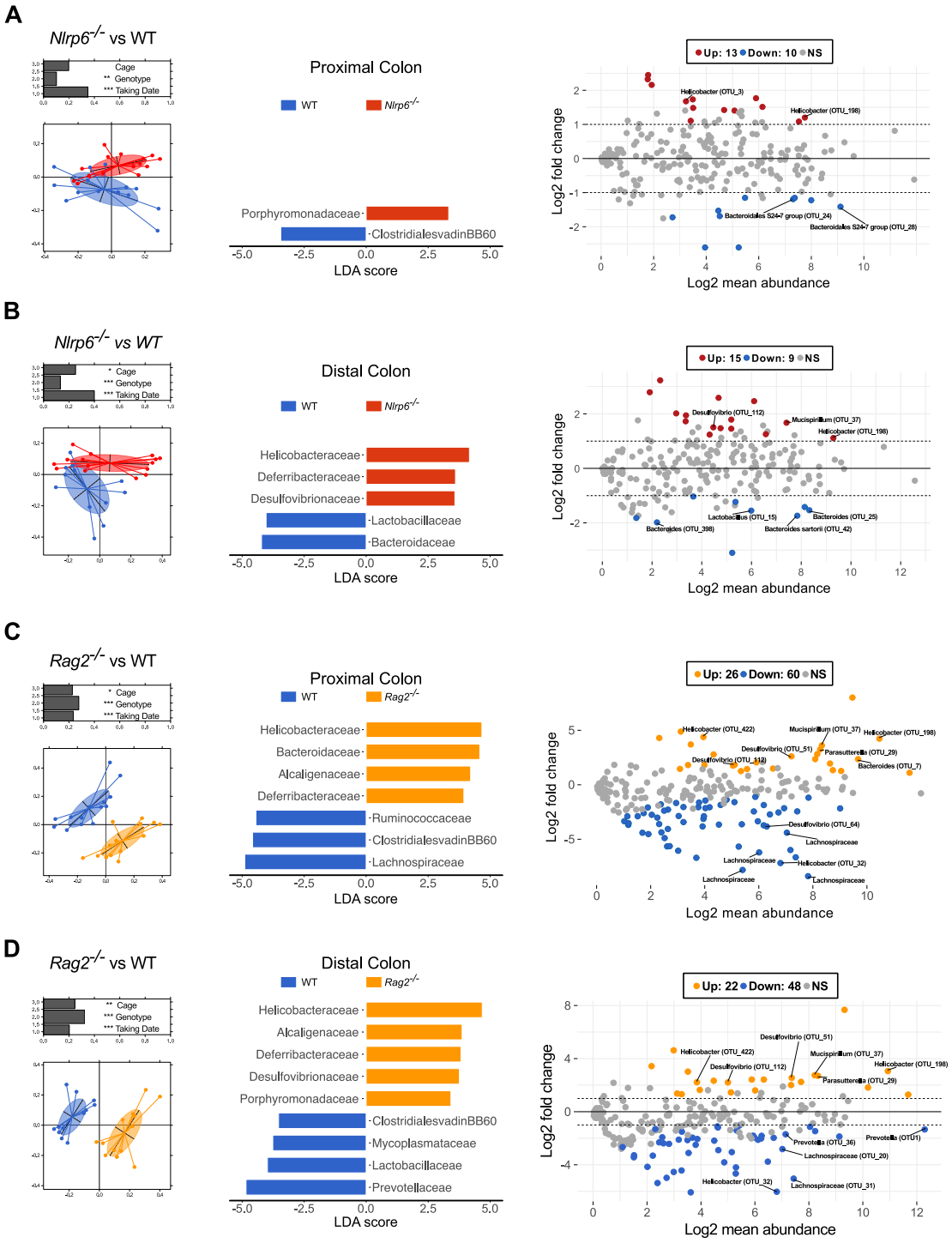

Figure 3.1.11: Increased abundance of distinct bacteria in distal colon of eSPF+ Conv-Dys *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice

(*Legend continued on next page*)

RNA-based microbiome analysis in the proximal (PC) and distal colon (DC) of two cohorts of eSPF WT, *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice 4 weeks after receiving independent FT from Conv-Dys *Nlrp6*$^{-/-}$ mice. (A-D) Active communities were analyzed using NMDS, LEfSe and DESeq2. After LEfSe analysis bacterial families with LDA > 3.0 are displayed. MA plots were used to visualize DA OTUs (Up: red; down: blue; P values of <0.05 after correction for multiple tests) identified using DESeq2. (A and B) Analysis of active communities in PC (A) and DC (B) of WT and *Nlrp6*$^{-/-}$ mice. (C and D) Analysis of active communities in PC (C) and DC (D) of WT and *Nlrp6*$^{-/-}$ mice. Data shown summarizes two independent experiments. Permutational multivariate analysis of variance (ADONIS) was used to calculate the variance explained by individual factors in (A and D). A significant effect was attributed when P-value is < 0.05 and R^2 is > 0.10 (equivalent to 10% of explained variance); ***P<0.001 **P<0.01, *P<0.05. Data represent two independent replicates (n = 14-15 mice per genotype). Adapted from [Galvez, E.J.C., *et al*., 2017]

Taking the results presented above, I investigated if certain microbial taxa could be associated with immune deficiencies. To identify if particular OTUs are associated to *Nlrp6*$^{-/-}$ as well as *Rag2*$^{-/-}$, I identify the fraction of shared OTUs from a total of 41 distinct OTUs enriched in gene-deficient mice and then the ones shared between the different colonic sites of *Rag2*$^{-/-}$ and *Nlrp6*$^{-/-}$ mice, respectively (Figure 3.1.12A). In total 9 OTUs were enriched in at least one site in both *Rag2*$^{-/-}$ and *Nlrp6*$^{-/-}$ mice, but only 2, OTU_198 and OTU_96 at all sites. The first OTU belongs to a member of the Helicobacteraceae (closest match *Helicobacter typhlonius* strain MIT 97-6810*,* 99% of identity, NCBI BLAST) and the second to a member of the Clostridiaceae (closest match *Butyricicoccus pullicaecorum* strain 25-3*,* 96% of identity, NCBI BLAST) (Figure 3.1.12B). Additional OTUs that were partially shared included *Mucispirillum schaedleri* OTU_37 (strain HRI I17, 100% of identity), *Desulfovibrio* OTU_112 (closest match *Desulfovibrio desulfuricans*, 90% of identity), and OTU_454 (unknown Lachnospiraceae UCG-001) (Figure 3.1.12C). Notably, *Prevotella*ceae did not have significant changes in their relative abundances (Figure 3.1.13). The relative low number of shared DA OTUs between gene-deficient mice (9 of 41) supports the model that specific OTUs are able to explore distinct niches opened as a consequence of impaired host immunity. But

notably the shared OTUs are enriched in potential pathobionts from the phylum proteobacteria suggesting that these bacteria potentially benefit from deficiencies in Nlrp6 and adaptive immunity.

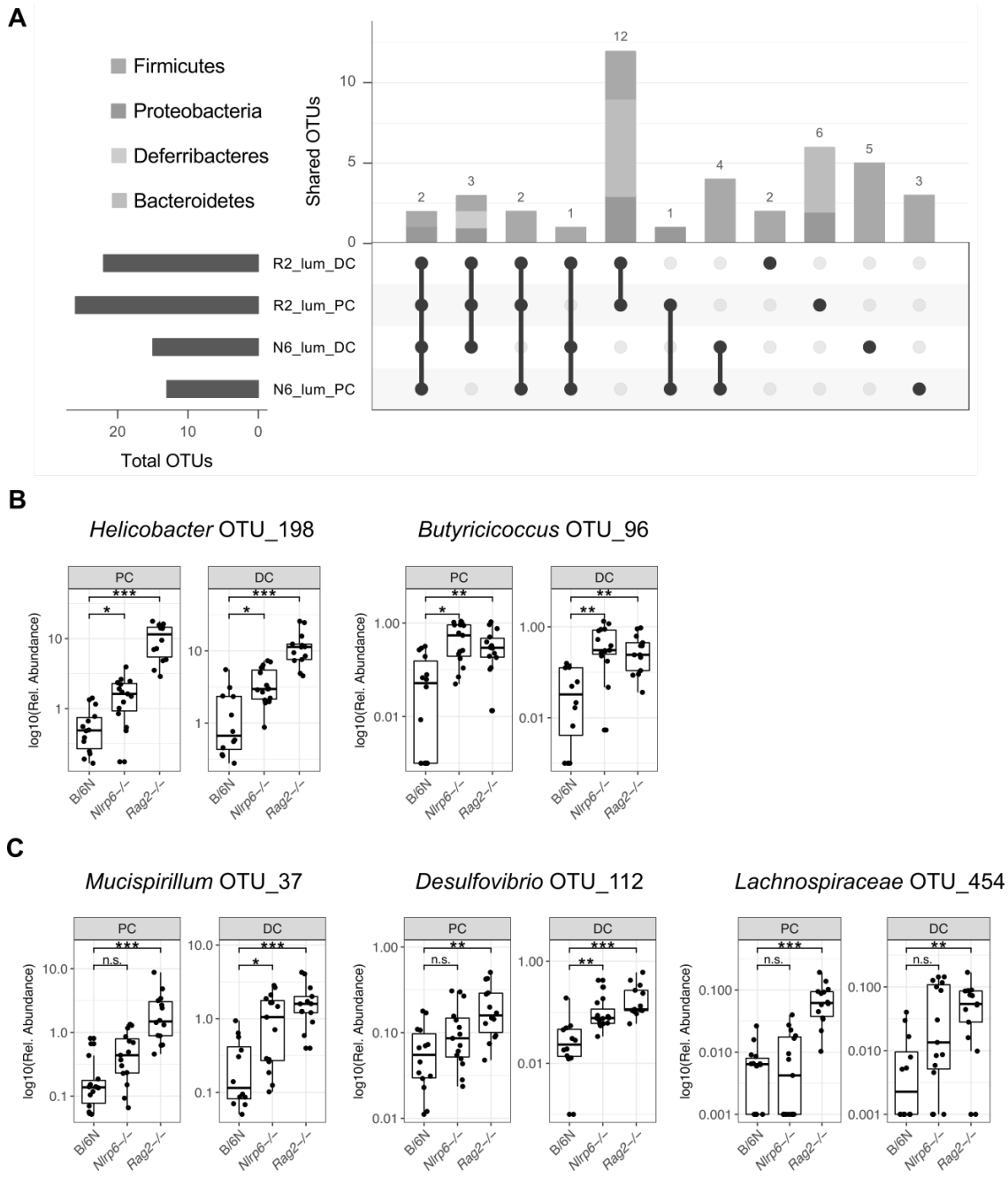

Figure 3.1.12: Distinct commensals explore specific niches in immunodeficient mice

RNA-based microbiome analysis in the proximal (PC) and distal colon (DC) of two cohorts of eSPF WT, *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice 4 weeks after receiving independent FT from Conv-Dys *Nlrp6*$^{-/-}$ mice. (A)

(*Legend continued on next page*)

Visualization of DA OTUs commonly enriched in *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice using UpSetR plot. The upper chart shows the number and taxonomy of shared OTUs that were identified in higher abundance in *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice. The matrix in the bottom visualizes the set of intersections represented by the connected dots. (B and C) Relative abundance of selected DA OTUs in the DC shared between *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice. P-values are from Mann-Whitney U test with Benjamini-Hochberg correction for multiple testing. Adapted from [Galvez, E.J.C., *et al*., 2017]

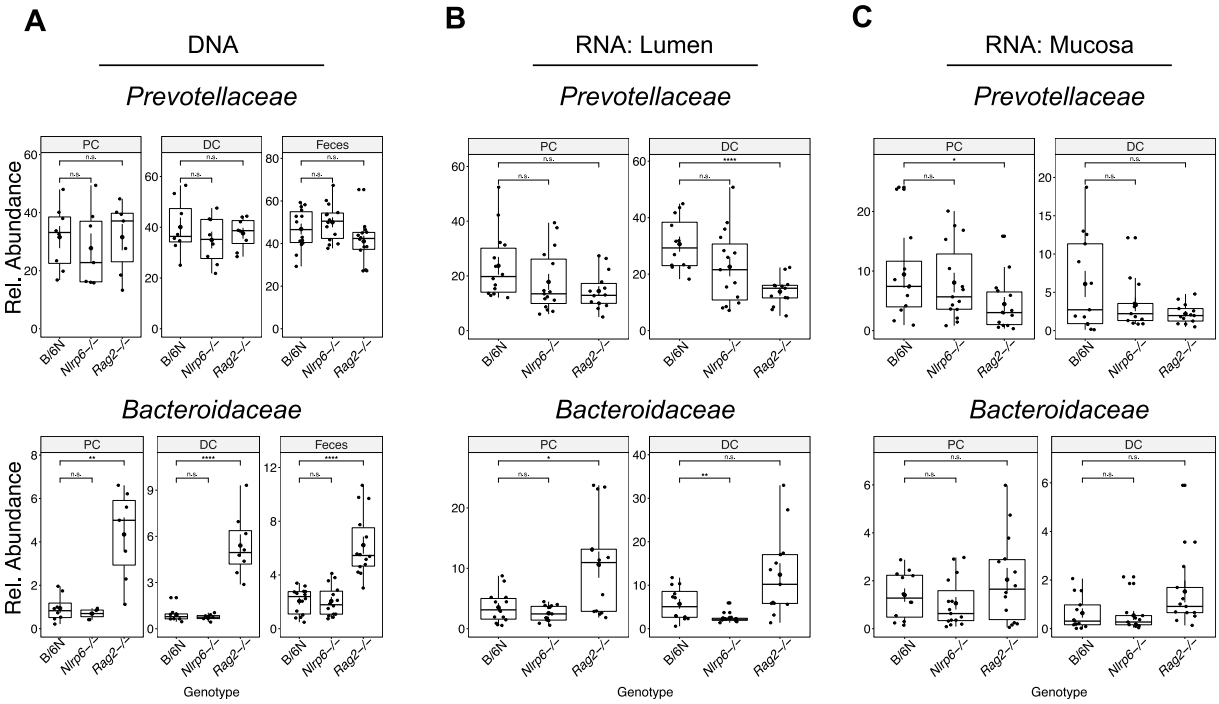

Figure 3.1.13: Unaltered abundance of *Prevotella*ceae and Bacteroidaceae in eSPF+Conv-Dys WT, *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice.

Microbiome and host immune analysis of WT, *Nlrp6*$^{-/-}$ and *Rag2*$^{-/-}$ mice 4 weeks after receiving independent FT from Conv-Dys *Nlrp6*$^{-/-}$ mice. (A) Relative abundances of *Prevotella*ceae and Bacteroidaceae using (A) DNA-based amplicons from feces and (B) RNA-based in lumen and (C) mucosa samples. Adapted from [Galvez, E.J.C., *et al*., 2017]

Together, our observation that the abundances of DA OTUs vary strongly at different anatomical sites reiterates the notion that specific immune mechanisms as well as environmental and ecological parameters shape microbial colonization throughout the gastrointestinal tract.

3.2 Diet and microbial competition modulates the prevalence of the diverse *Prevotella* genus in the mouse gut

The gut microbiome is typically dominated by bacterial species belonging to the phyla Bacteroidetes and Firmicutes, while members of the phyla Proteobacteria, Actinobacteria, Verrucomicrobia and Fusobacteria are found in lower abundances (*3*). Despite the predominance of Bacteroidetes and Firmicutes in the gastrointestinal tract of humans, large variations of the microbiome composition, specifically on the level of prevalent species (*3*, *111*) and gene functionality (*112–114*) have been reported between individuals. Based on an initial survey of 33 healthy patients from Europe, Japan, and North America, in 2011 Arumugam *et al.,* proposed that their microbiota could be clustered in three distinct bacterial profiles or “enterotypes”. These clusters or densely populated areas in an ordination space displayed a similar composition pattern with subjects dominated by one of three different bacterial genera: *Bacteroides*, *Prevotella*, or members of the order Clostridiales, frequently governed by the genus *Ruminococcus*. Whether the assembly of the human gut microbiome follows distinct rules or it is a consequence of stochastic colonization events regulated by host genetics and environmental factors, is a field of high controversy and intense discussion (*19–22*). Hence, more studies are needed in order to apply the concept of “enterotypes” in personalized medicine with the aim to stratify patients for better diagnosis and treatments. Since the introduction of the enterotypes, the understanding of the biology

of *Bacteroides*, *Prevotella* and *Ruminococcus* have largely increased, especially the abundant members of the Bacteroidaceae family such as *B. thetaiotaomicron*, *B. ovatus,*, *B. fragilis*, and *B. vulgatus* among others, which have been successfully isolated, characterized and genetically manipulated (*37*, *115*, *116*). Yet, while the functional description of *Bacteroides* spp*.* continues to being expanded (*117*), the diversity and functional niche of *Prevotella* and many members of Clostridiales remain elusive due to the low cultivability and the lack of methods for genetic engineering of these microbes.

Nevertheless, *Prevotella* species are known as a group of anaerobic Gram-negative bacteria of the Bacteroidetes phylum widely spread in distinct body habitats such as, oral cavity, urogenital and the gastrointestinal tract, in which they have been suggested to play a key role in the degradation of complex polysaccharides (*118*). In the human gut, *Prevotella copri* represents the most abundant species followed by *Prevotella stercorea* (*119*). Recent studies have associated the prevalence of *Prevotella* to non-Westerners who consume a plant-rich diet (*43*, *44*). Furthermore, it has been shown that *Prevotella copri* can improve glucose metabolisms stimulated by the intake of prebiotics (*120*). Together, these studies suggest that *Prevotella* is a beneficial commensal. In contrast, other studies have associated *Prevotella* spp*.,* with autoimmune diseases and gut inflammation. Specifically, an overabundance of specific subtypes of *Prevotella copri* were associated with increased risk for rheumatoid arthritis (*121*). In mouse models, an altered gut microbiota dominated by an uncultured member of the genus *Prevotella* was discovered in NLRP6-deficient mice and responsible for higher susceptibility to chemically-induced colitis (Elinav et al., 2011). In addition, colonization by intestinal *Prevotella* in mice results in metabolic changes in the intestine,

which was associated with reduction in interleukin (IL)-18 production and consequent exacerbation of the intestinal inflammation (Iljazovic, A., unpublished observation). These data suggest that a *Prevotella*-dominated microbiome may have the propensity to promote inflammation and intestinal dysbiosis. The conflicting role of *Prevotella* on the host physiology may be explained by the high species diversity and their functional potential. However, the investigation of these hypotheses is prohibited by the poor characterization of the *Prevotella* ecology, their functional niche, and the unknown species and strain diversity.

In the present work I characterized the taxonomic diversity and genomic potential of three uncharacterized *Prevotella* species; subsequently I investigated the interspecies fitness (ability to replicate and survive in a competitive environment) and their metabolic niche. Specifically, in order to understand the *in vivo* functional role of *Prevotella,* we first isolated four prevalent *Prevotella* species from mice intestine and sequenced and assembled their genomes. Furthermore, I assessed the genomic potential through interspecies competition experiments and finally I performed RNA-seq of isolated *Prevotella* species in two defined intestinal communities. In summary, I unveiled surprisingly high genome variability and functional responses between three new species of the genus *Prevotella* and identified a set of Polysaccharide Utilization Loci (PULs), which may play a key role in the fitness and inter-species competition in the intestinal ecosystem.

3.2.1 Distinct *Prevotella* OTUs govern the mouse gut microbiome

A variable abundance of members of the genus *Prevotella* has been observed in the human gut microbiome and it has been hypothesized that strain diversity could be responsible for contradictory phenotypes in health and disease (*46*, *123*, *124*). Similar to humans, we previously identified distinct *Prevotella* taxonomic units (OTUs) (n=5) (*125*), whose abundance differed strongly between lines of mice maintained within the same facility. With the aim to comprehensively characterize the taxonomic diversity of intestinal *Prevotella* spp*.,* in mice, we screened the intestinal microbiota composition of twelve mouse lines from diverse sources for the presence of *Prevotella* spp. Specifically, I included three mouse lines maintained at the HZI, i.e. eSPF (*125*), N6_dys (*47*), and NCI1090 (*126*), as well as nine mouse lines from distinct hygiene barriers of four different vendors of laboratory mice (Ch_rivers, Taconic, Harlan, Janvier) (mice n = 4–10, barriers n = 1-3/vendor) (Figure 3.2.1A). The microbiota composition and the relative abundance of *Prevotella* were characterized by sequencing of the V4 region of the 16S rRNA gene followed by phylogenetic inference (see Materials and methods). PCoA ordination analysis using Bray-Curtis distances demonstrated distinct clusters by vendor, while barrier had a smaller effect within each vendor. Then, I performed permutational multivariate analysis of variance (ADONIS) to estimate the size effect of the variables “Vendor” and “Barrier” on microbiota composition. Our results revealed that the largest component of microbial variation was attributed to the variable “Vendor”, which explained 41% of the samples variability (ADONIS, R^2=0.41 P-value < 1e-04) (Figure 3.2.1B). In line with previous reports (*127*–*129*), the microbiota in the analyzed mouse lines are predominantly composed of

members belonging to the phyla Firmicutes, Bacteroidetes, Verrucomicrobia and Proteobacteria (Figure 3.2.1D). With the aim to understand the samples grouping based on the abundance of *Prevotella*, I calculated the *Prevotella* ratio (abundance of *Prevotella*/sum(Bacteroidales)) for each sample. When I colored the PCoA by the *Prevotella* ratio, I observed an abundance gradient that correlates with the vendor's clusters (Figure 3.2.1C). Interestingly, the *Prevotella* ratio differed significantly among the vendors and barriers (P-value <0.001, Kruskal-Wallis) (Figure 3.2.1E). Two of the mouse lines maintained at the HZI were characterized by the highest *Prevotella* ratio (N6_dys and NCI-1090) followed by the vendors Ch_rivers and Janvier.

Analysis of members of the *Prevotella* genus at the OTU level (USEARCH, similarity 97% within V4 region, abundance > 0.02%), revealed 5 distinct *Prevotella* OTUs that dominate each vendor's gut community (Figure 3.2.1F). Of all detected *Prevotella* phylotypes, OTU_1 was the most prevalent being present in the vendors Janvier (mean= 14.48%), Ch_river (mean= 12.84%) and Harlan (mean = 7.70%). The other *Prevotella* OTUs were largely specific to unique vendors and mouse lines, respectively. Specifically, OTU_12 was the highest abundant in NCI-1090 mice (mean= 27.37%), followed by OTU_16 from N6_dys (mean= 24.04%), OTU_15 in Ch_river (mean= 7.38%) and OTU_26 observed in Taconic and N6_dys (mean= 3.00% and 0.89%) respectively. Phylogenetic analysis of the total fraction of Bacteroidetes showed that *Prevotella* OTUs are grouped in three main lineages and the large sequence distance between groups suggested that each OTU could be associated to distinct species (Figure 3.2.1G).

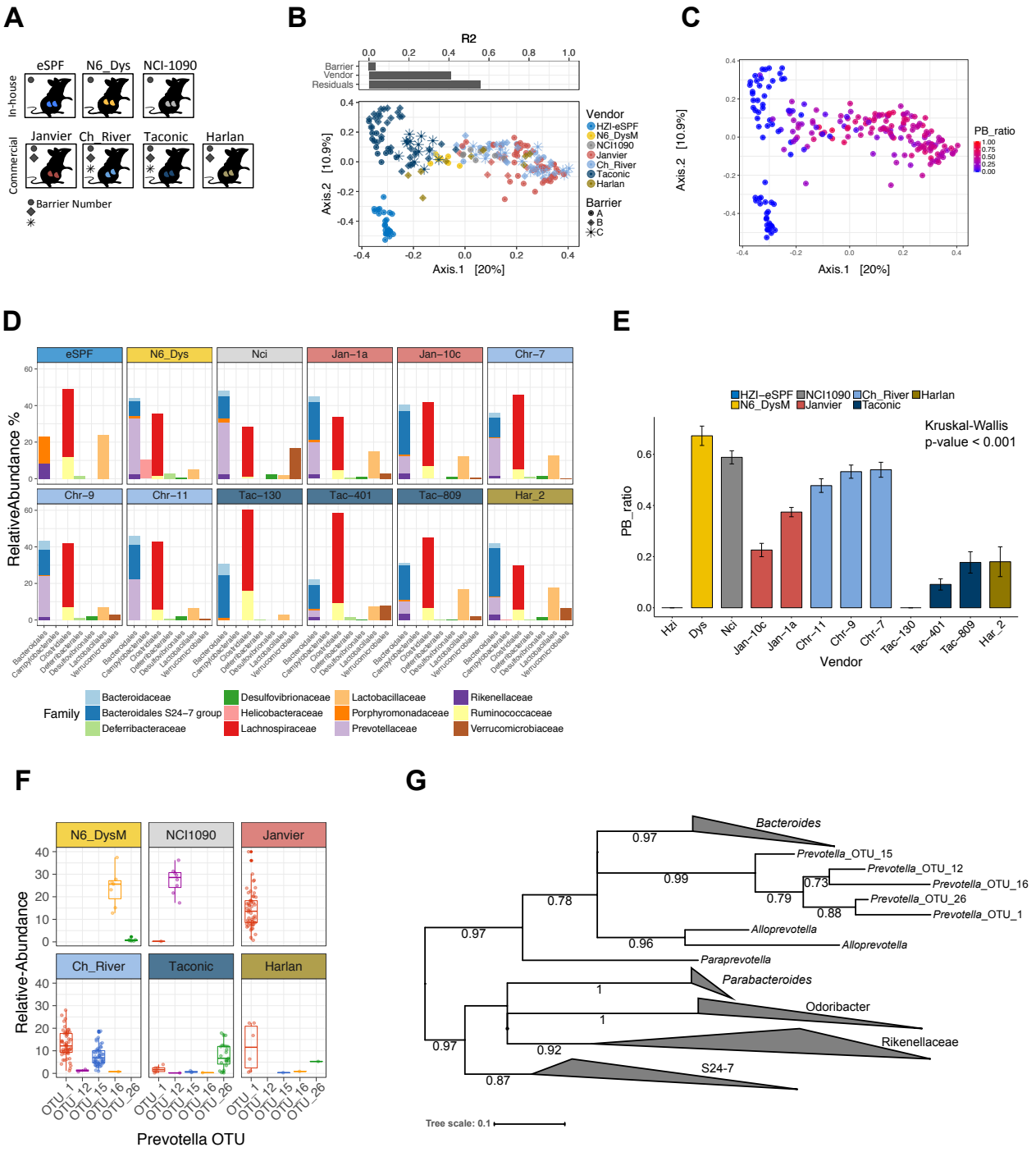

Figure 3.2.1: Identification of distinct *Prevotella* OTUs in the mouse gut microbiome

(A). Overview of murine vendors screened for *Prevotella*, boxes indicates mice vendors; inner shapes indicate the number of screened barriers. (B) Variance effect size using ADONIS test and Principal component analysis (PCoA) of microbiota composition using Bray-Curtis distances. (C) PCoA showing samples colored by the *Prevotella* ratio (relative abundance of *Prevotella*/sum(Bacteroidales), red indicates high *Prevotella* and blue indicates absence of *Prevotella*. (D) Microbial signatures per barrier in

(*Legend continued on next page*)

each mouse vendor; X-axis indicates bacterial composition at order level, Y-Axis indicates the mean relative abundance per barrier and bar colors represent bacterial family. (E) Relative abundance ratio between *Prevotella* to the total of Bacteroidales, P-value <0.001 Kruskal-Wallis. (E) Boxplot distribution of relative abundance of identified *Prevotella* OTUs (clustering similarity > 97%), n = 7–60. (F) Phylogenetic tree of Bacteroidales OTU fraction using RaxML, the numbers above the branches are support values when bootstrapping is larger than (0.70) equivalent to 70%.

3.2.2 Isolation of prevalent *Prevotella* spp. from the mouse gut

Prevotella is a prevalent member of the gut microbiome in mammals, but until now no *Prevotella* spp. have been isolated and taxonomically described from the mouse intestine. The analysis based on OTUs suggested the presence of five distinct *Prevotella* species prevalent in mice from different vendors. Hence, we used a targeted culturing strategy from the intestinal content of the respective mouse lines. Importantly, this approach included a PCR-based screening for members of the genus *Prevotella* to enable screening of a large number of individual colonies. A total of four species of the genus *Prevotella* were isolated (Aida Iljazovic, Doctoral thesis TU Braunschweig) with distinct morphologies (Figure 3.2.2A). Then, I performed a phylogenetic analysis based on the full-length 16S rRNA gene sequences. Interestingly, the analysis revealed a similarity of 100% with the obtained OTUs from the vendors Janvier, NCI, N6_dys and Taconic (Figure 3.2.2B).

Next, with the aim to determine the prevalence and relative abundance of the novel *Prevotella* isolates globally within microbial communities of diverse sets of hosts, I employed the IMNGS tool (www.imngs.org), which analyses the entire Sequence Read Archive (SRA-NCBI) for the presence of specific 16S rRNA gene sequences (for details see Methods). First, I calculated the prevalence (the percentage of samples where a given OTU is detected) using two sequence similarity thresholds (99, 97%). The results showed a wide distribution of murine *Prevotella* spp. when I compared six distinct hosts

(primates, bovines, rat, mouse, pig and humans) (Figure 3.2.2C). At 97% of similarity (Considered as “species” level), the OTU_1 presented a wide distribution in multiple hosts, in rodents (mouse 12.41% of 18,122 and rat = 45.21% of 637 samples), in bovines (8.78% of 899) and primates (31.30% of 115), followed by OTU_26, which was detected mainly in the gut of rodents (mouse 25.52% of 18,122 and Rat = 10.83% of 637) but not in primates. The OTU_12 was found in mouse samples (2.94% of 18,122) and bovine (0.22% of 899), while OTU_16 was found mainly in the mouse gut (6.67% of 18,122). Of the four species, three were rodent-specific, while OTU_1 was widespread in distinct hosts. These results potentially reflect the diverse metabolic capacity of the intestinal *Prevotella* spp., and indicate that the novel isolates represent prevalent species that have been frequently observed in distinct studies.

Then, I selected the SRA datasets that were positive for *Prevotella* OTUs (similarity > 97% and abundance > 0.1%) and I determined the mean relative abundance for each host (Figure 3.2.2D). Interestingly, I found that the OTU_26 is a common member of the mouse gut microbiome but in a relative low abundance (n= 350, rel. abundance mean = 1.59%, SD= 2.30), while OTU_1 is found in samples from mouse, rat and primate where it constitutes more than 3% of the relative abundance (Mouse: n=125, mean=3.92%, SD=4.62; Rat: n=24, mean=3.94%, SD=3.35, Primate: n=3, mean=7.11%, SD=2.31). The OTU_16 and OTU_12 were found specifically in mouse samples, with OTU_16 being the most abundant of the four (OTU_16: n=52, mean=13.34%, SD=13.34; OTU_12: n=3, mean=1.60%, SD=0.5, respectively).

Based on host specificity and relative abundance, we suggest the following names for the novel isolates: OTU_1 = *Prevotella rodentium sp. nov.,* OTU_12 = *Prevotella muris sp. nov.,* OTU_16 = *Prevotella intestinalis sp. nov.*, OTU_26 = *Prevotella musculis sp.*

nov. These names have been implemented in the next sections of the thesis. Of the four strains, three were successfully regrown after cryopreservation and were deposited at the German Collection of Microorganisms and Cell Cultures.

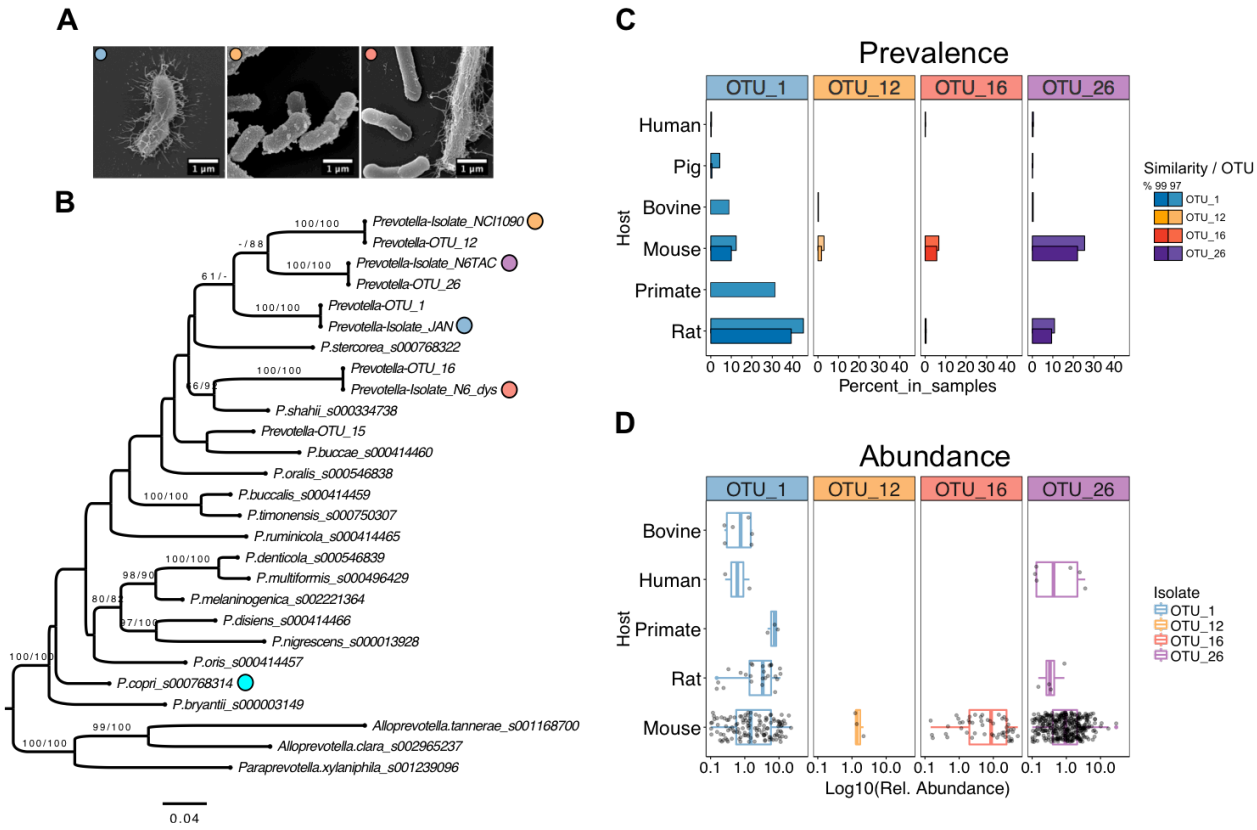

Figure 3.2.2: Phylogeny and prevalence of novel *Prevotella* species

(A) Scanning electron microscopy (SEM) of cultivable *Prevotella* species, cells at ×35,000 magnification. Scale bar represents 1 µm. Colored dots indicate OTU source (blue= Janvier, orange=NCI1090, red=N6_Dys). (B) Phylogenetic analysis of isolates vs OTU representative 16S rRNA gene sequences. The ML tree was inferred under the GTR+GAMMA model and rooted by midpoint-rooting. The branches are scaled in terms of the expected number of substitutions per site. The numbers above the branches are support values when larger than 60% from ML (left) and MP (right) bootstrapping. (C) *Prevotella* isolates prevalence and (D) average relative abundance in SRA datasets for 16S rRNA gene amplicon sequences (SRA-NCBI). SEM was performed by Manfred Rohde (Panel A).

3.2.3 Comparative genome analysis unveils high diversity of intestinal *Prevotella* spp.

Recently, a high divergence in the gene repertoire among the *Prevotella*ceae family from distinct body sites has been reported (*130*), and others studies have suggested a high OTUs diversity of human intestinal *Prevotella* spp. observed by 16S rRNA gene amplicons (*131*). In order to compare the genomes and functional diversity of the uncharacterized *Prevotella* isolates, I performed whole genome sequencing using the Illumina HiSeq technology. After assembly of reads, I obtained four draft genomes with a mean length of 3.3 Mbp and 46.5% GC content. The largest genome observed was *P. rodentium* with 3.539 Mbp, followed by *P. intestinalis* and *P. muris* (3.516Mbp and 3.355 Mbp, respectively), the smallest genomes belonged to *Prevotella musculis* with 3.003 Mbp (Table 3.2.1). Then, I compared the genomic features using three different approaches. First, I analyzed the genome synteny and nucleotide similarity by whole genome alignment using as reference the genomes from *Prevotella copri* and *Prevotella ruminicola,* both abundant in the human intestine, as well as the closest-matching reference strain *(Prevotella shahii)* as identified previously by phylogenetic analysis of the 16S rRNA genes (Figure 3.2.3B). The results revealed unexpected low similarity with several genome rearrangements (Figure 3.2.3A). Identification of the conserved regions corresponded to ribosomal genes (16S, 23S rRNA genes and tRNAs), transposable elements and prophages. Second, I performed pangenome analysis using a protein clustering approach (cd-hit, protein similarity > 50%) and functional analysis by predicting and comparing the protein annotations using the SEED tool (*72*, *73*). A mean of 2909 features was predicted for the *Prevotella* isolates with 72.1 % of the CDS assigned to a known gene or SEEDs cluster and 28% not being

annotated or associated with a functional cluster. Therefore, these were classified as hypothetical genes. From the total fraction, only 33.4% of the features were associated with a functional role or SEEDs subsystem category (Figure 3.2.2B, left panel). I observed a correlation between the numbers of CDS with the genome sizes. A total of 2916 CDS were predicted for *P. intestinalis,* followed by *P. rodentium* and *P. muris* (2881 and 2811 CDS, respectively), the smallest number corresponded to *P. musculis* with 2797 CDS. Comparison of the fraction of identified genes resulted in 775 shared features between the sequenced isolates, accounting for 30% of all CDS. Interestingly, *P. intestinalis* presented the highest number of unique features (n=1683) and did not share a set of 358 CDS presented by the others strains (Figure 3.2.2B, right panel).

Table 3.2.1 Genome assembly statistics *Prevotella* isolates

Name	*P. rodentium*	*P. intestinalis*	*P. muris*	*P. musculus*
OTU ID	OTU_1	OTU_12	OTU_16	OTU_26
Mean rel. abundance	14.48%	27.37%	24.04%	3.00%
Size in bp	3539	3516	3355	3003
GC content	47.7	47.0	44.4	47.2
Number of contigs	68	18	26	35
Number of coding sequences	2916	2881	2811	2797
Number of RNAs	54	62	64	53

Shared features by all species included genes involved in anaerobic degradation of simple carbohydrates and biosynthesis of amino acid pathways. In contrast, the identification of the unique features in *P. intestinalis* showed potential virulence factors such as metalloproteases (e.g. peptidase M6), cysteine-type peptidase and a large set of uncharacterized proteins associated to polysaccharide utilization loci (PUL) systems. Finally, I performed functional category analyses using the SEEDs subsystem

annotations. Fraction of 33.4% of the CDS were assigned to a known category and I observed that all four *Prevotella* species encode in their genomes a high number of genes for amino acids, carbohydrates and protein metabolism (Figure 3.2.3C). Analysis at the subcategory level (SEEDs subsystem) unveils the presence of several copies of Bacteroidales transposons and ribosomal proteins. Of the amino acid metabolism, the most abundant genes were those involved in methionine biosynthesis as well as glutamine, glutamate, aspartate and asparagine. The carbohydrate fraction was represented by Xylose, Lactose and Galactose utilization systems.

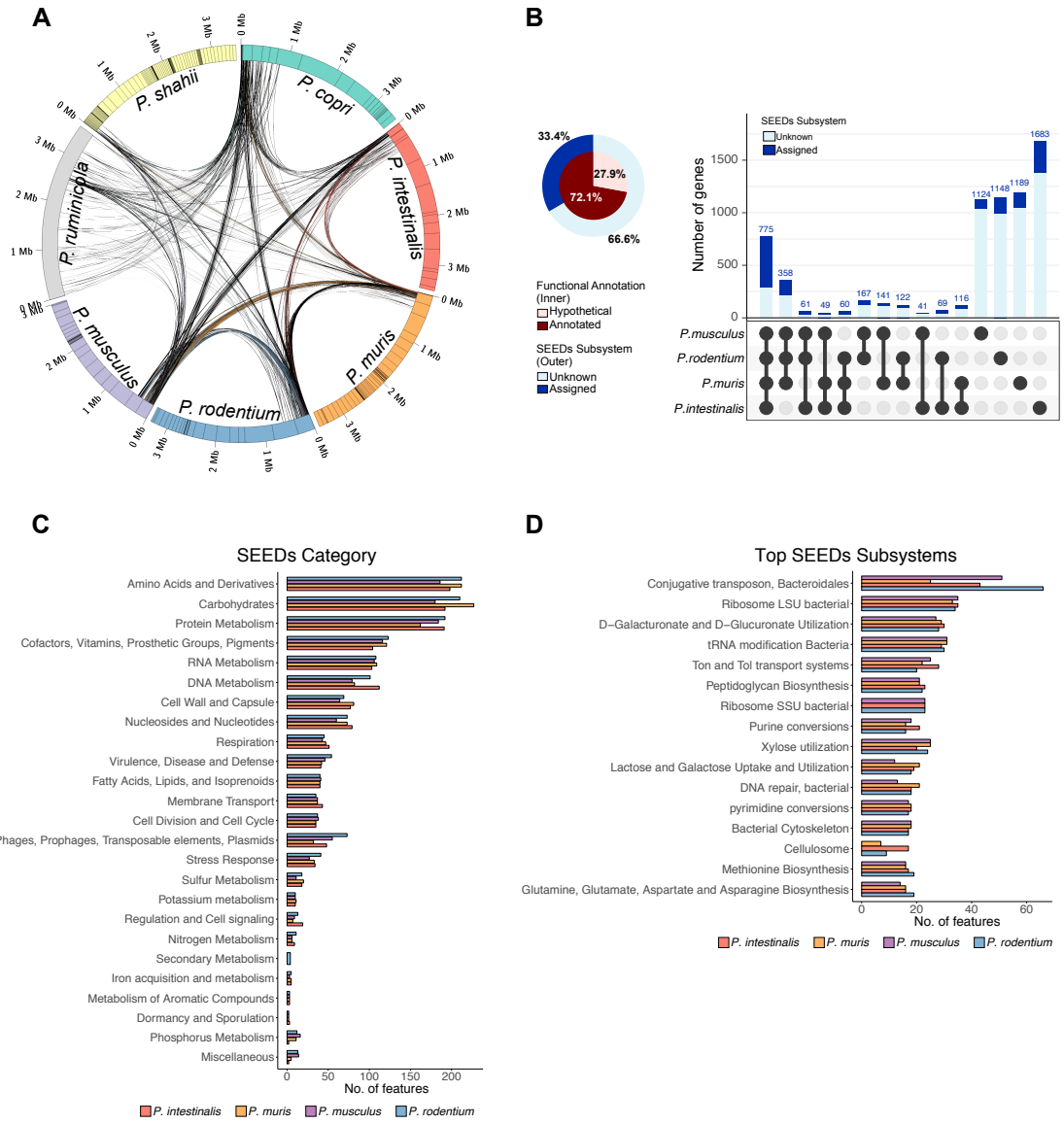

Figure 3.2.3: wide-genome analysis and functional characterization of *Prevotella* spp.

(A) Circular representation of genome assembly and nucleotide comparison vs intestinal *Prevotella* reference strains. Outer bars represent each bacterial chromosome, segments indicate contig's' length and the inner connections indicate a genome fragment alignment. (B) Functional genome comparison. Genes were predicted and annotated using the SEED subsystems and the RAST tool. Left panel a pie chart indicating the fraction of annotated CDS, inner indicates fraction of annotated genes, outer ring shows the proportion of CDS associated with a SEED subsystem category. (C) Distribution of SEED functional categories. (D) Top list of SEED subsystems (Subsystem > 20 CDS).

3.2.4 Competition dominates interactions among *Prevotella* spp.

To compare the metabolic fitness associated with the variable genomic potential of each of the *Prevotella* spp., we performed a competition experiment mixing the species followed by quantification of the colonization dynamics. To this end, we cohoused five different mouse lines: three mouse lines harboring a distinct *Prevotella* species, (N6_DysM (*Nlrp6*$^{-/-}$), Janvier (wt), NCI1090 (wt) and two mice lines lacking members of *Prevotella* (eSPF (wt) and an out-bred germ-free strain (exGF SwissWebster)) (Figure 3.2.3A). Five mice, i.e. one from each line, were co-housed together for 3 weeks, and this setup was replicated four times. We collected fecal pellets before co-housing, after 1 week and finally after the 3-week period. After co-housing, mice were sacrificed and DNA was extracted from the distal colon content. The relative abundance was estimated using 16S rRNA gene sequencing. Ordination analysis considering the variables “Microbiota” and “Time point” unveiled distinct composition dynamics over time (Figure 3.2.3B). Samples before cohousing (week0) showed specific clusters by mouse line (highlighted in colors). Interestingly, after one week of cohousing the samples followed two different fates, the exGF and NCI1090 mice switched their microbiota closely to the N6_dysM samples, while the eSPF samples were more similar with Janvier samples. Subsequently, after three weeks of cohousing all samples converged in one cluster close to the native Janvier microbiota.

Alpha diversity analysis revealed that before cohousing the Janvier microbiota had the highest richness and diversity (Richness = 82.72, SEM= 4.53; Shannon = 3.17, SEM= 0.07, P = 0.0145; Kruskal-Wallis test), followed by eSPF (Richness = 58.06, SEM= 5.19; Shannon = 2.87, SEM= 0.22), N6Dys (Richness = 39.40, SEM= 2.73; Shannon = 2.43,

SEM= 0.09) and NCI1090 (Richness = 26.50, SEM= 1.68; Shannon = 1.93, SEM= 0.05) (Figure 3.2.3C). During cohousing the alpha diversity of the fecal communities in all mouse lines besides the Janvier mice increased during the following three weeks after no differences in OTU richness and diversity remained (Mean richness at week3 = 74.78, SEM= 3.01; Shannon = 2.92, SEM= 0.06).

Taxonomic analysis of microbiota composition before co-housing (Week0) confirmed the presence of the three *Prevotella* species in the respective mouse lines and the absence in the eSPF microbiota (Figure 3.2.3D). The overall taxonomy profile at family level was mainly governed by OTUs of the families Lachnospiraceae and *Prevotella*ceae as well as the little characterized Bacteroidales S24-7 group. After 1-week cohousing, I observed distinct colonization patterns according to each mouse line; the exGF mice are highly colonized by Bacteroidaceae while *Prevotella*ceae readily colonized eSPF mice. After the 3-week cohousing, the microbiome composition of all samples presented an even composition with a high relative abundance of *Prevotella*ceae. These observations are in line with the above-mentioned ordination and alpha diversity results.

Next, I wanted to identify whether the over-expansion of *Prevotella*ceae in the co-housing experiment was a consequence of cooperation or rather reflect competition between the three *Prevotella* species. Using the 16*S* rRNA gene reference sequences from the novel *Prevotella* genomes I identify the relative abundance for each of the species. Interestingly, I observed an overexpansion of the species *P. intestinalis* after the first week of cohousing in eSPF mice reaching up to 36.45% (SEM= 6.52%) of the relative abundance. The second most abundant *Prevotella* species in the first week of cohousing was *P. rodentium*, which interestingly rapidly colonized the exGF mice (mean

rel. abundance = 3.44%, SEM= 2.03%). *P. muris* displayed the lowest colonization and its relative abundance dropped from 14.75% (SEM= 2.78) before co-housing to 5.57% (3.48) in the 1-week in NCI1090. In the third week, *P. intestinalis* outcompeted *P. rodentium* and *P. muris* to reach in all mouse lines a high relative abundance (mean rel. abundance = 30.39%, SEM= 2.52) in contrast to *P. rodentium* (mean rel. abundance = 1.28%, SEM= 0.34) and *P. muris* (mean rel. abundance = 0.52%, SEM= 0.11). Combined, these results indicated that *Prevotella* species strongly compete in the mouse gut and the high genome diversity does not reflect complementary metabolic functions for cooperation as observed in some species of *Bacteroides* (*132*).

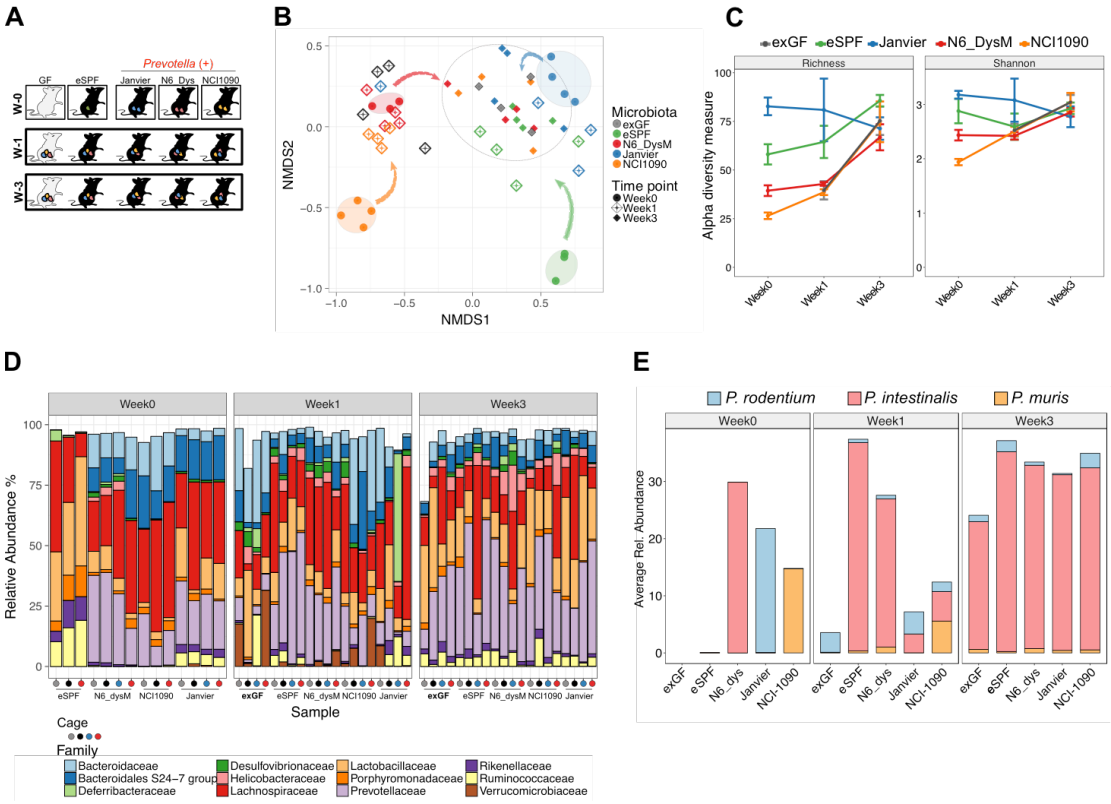

Figure 3.2.4: *In vivo* competition experiment reveals predominance by one single *Prevotella* specie

(A) Scheme of competition experiment. (B) NMDS ordination analysis (Bray-Curtis distances) of microbiota composition before (week0) and during cohousing (Week1, Week3). (C) Estimation of α-diversity using Richness and Inverse Simpson index. (D) Relative abundance at Family level before and during cohousing. Labels indicated the group of mice before cohousing (n=4) and colored dots represent each cage. (E) 16S rRNA relative abundance showing *Prevotella* species dynamics.

3.2.5 Functional signatures of *Prevotella* isolates reveal degradation of complex carbohydrates and nitrogen utilization

Next, I wanted to identify *in vivo* correlates in the gene expression profiles of the different *Prevotella* species to eventually understand how *P. intestinalis* outcompetes *P. rodentium* and *P. muris in vivo.* Therefore, I performed RNA-seq for each isolate in WT C57BL/6N mice (*in vivo*) and during growth under laboratory conditions (*in vitro*). Since the microbiota composition differed in each of the mice vendors, the transcriptomes *in*

vivo were performed by colonization of each *Prevotella* species in the standardized mouse model *eSPF* (*Prevotella* free) as well as in the original mouse vendor (*Complex*) from which each *Prevotella* specie was isolated.

Based on our previous gut biogeography characterization, which identified the highest abundance of *Prevotella* OTUs in the proximal colon, high-quality total RNA (RIN > 8) was isolated from the luminal content at this site for subsequent *in vivo* transcriptome analysis. For the *in vitro* gene expression profile, each isolate was grown anaerobically in (BHI) supplemented with 10% FBS and 0.5 g/L vitamin K (BHI-S+) at 37ºC until around OD=0.2 when total RNA was isolated. Then, for each experiment I performed ribosomal RNA depletion followed by cDNA synthesis and strand-specific Illumina-based sequencing (n= 4 biological replicates).

A total of 60.7 Gb (Single end reads; length 50 bp) were obtained for all samples (n=36), after quality filtering and uniquely mapping against the novel assembled *Prevotella* genomes I retrieved an average of 10.3 millions of reads mapping against the *Prevotella* genomes per sample (min= 892.554, max= 42.191.861). With the aim to identify the gene expression signatures of the *Prevotella* isolates, I first compared transcriptome analyses of *in vivo* (eSPF+ and Complex) vs *in vitro* samples for each of the *Prevotella* species (Figure 3.2.4-3.5). As expected, the gene expression profiles showed a distinct transcriptome profile for each treatment and the four biological replicates clustered together by experiment. Principal component analyses revealed that at least 53% of the variance is explained by the first component (PC1: *in vitro* vs *in vivo*) in *P. rodentium*, *P. intestinalis* (72%) and *P. muris* (74%). Interestingly, I observed an unexpected large variation between the transcriptional profiles *in vivo* (eSPF+ vs

Complex) demonstrating the impact of microbial composition on shaping the transcriptional repertoire in *Prevotella* species (Figure 3.2.5A).

To explicitly unravel the transcriptional responses *in vivo*, I first compared samples from colonized eSPF+ vs *in vitro* and then I compared the transcriptional alterations due to microbial composition (Complex vs. eSPF+). Using DESeq2 package (*76*) with a strict cutoff (Log Fold change > 2 and P-adjusted value < 0.01) I identified the set of differentially expressed genes (DEGs) in eSPF+ as well as in complex microbiota.

3.2.5 The metabolic niche of *P. intestinalis*

I identified a total of 752 DEGs (376 up and 376 down) in the distal colon of eSPF+ mice vs *in vitro*. From the total set of DEGs, a fraction of 35.1% corresponded to non-characterized CDS ("hypothetical protein" = 264, annotated = 488 CDS) (Figure 3.2.5B). In complex microbiota (N6_Dys) vs eSPF+, I identified a total of 403 DEGs (157 up and 246 down) with a fraction of 38.2% of hypothetical CDS ("hypothetical protein" = 154, annotated = 249) (Figure 3.2.5C).

The annotated set of highly expressed genes *in vitro* corresponded mainly to central metabolism such as degradation of simple sugars (e.g., D-glucose *glk* [EC:2.7.1.2], galactose metabolism, *galA* [EC:3.2.1.22]) and synthesis of amino acid (*ansA* [EC:3.5.1.1]). Moreover, several genes involved in thiamin metabolism (*thiC* [EC:4.1.99.17], *thiD* [EC:2.7.1.49, 2.7.4.7], *thiE* [EC:2.5.1.3], *thiF*, *thiG* [EC:2.8.1.10], *thiH* [EC:4.1.99.19], *thiS*) and tellurium stress response (*TerA* and *TerD*) were strongly up-regulated (Figure 3.2.5.D).

The *in vivo* eSPF+ transcriptome of *P. intestinalis* comprehended several genes involved in the degradation of complex glycans, especially multiple copies enzymes

associated with xylan degradation (endo-1,4-beta-xylanase [EC:3.2.1.8], *xylA* [EC:5.3.1.5], *xylS* [EC:3.2.1.177]). This group of genes was also found highly induced in "complex microbiota" condition and they constitute the shared-set of gene signatures up-regulated *in vivo* vs *in vitro*. Interestingly, the most strongly up-regulated genes in eSPF+ were those implicated in the high-affinity ammonia system (*glnA* [EC:6.3.1.2], *gltB* [EC:1.4.1.13, 1.4.1.14], *gltD* [EC:1.4.1.13, 1.4.1.14], *gdhA* [EC:1.4.1.4], *amtB*, *asnA* [EC:6.3.1.1], L2CL-diaminopimelate aminotransferase [EC:2.6.1.83], *dapF* [EC:5.1.1.7], *purF* [EC:2.4.2.14]). In contrast, differential gene expression in complex microbiota (N6_Dys) was not associated with a particular metabolic pathway and the top of overexpressed genes are part of wide cellular functions like processing of genetic information (e.g. transcriptional regulator *GntR*, conjugative transposons), tetracycline antibiotic resistance (*TetQ*) and iron acquisition systems such as the periplasmic siderophore binding protein (Figure 3.2.5D).

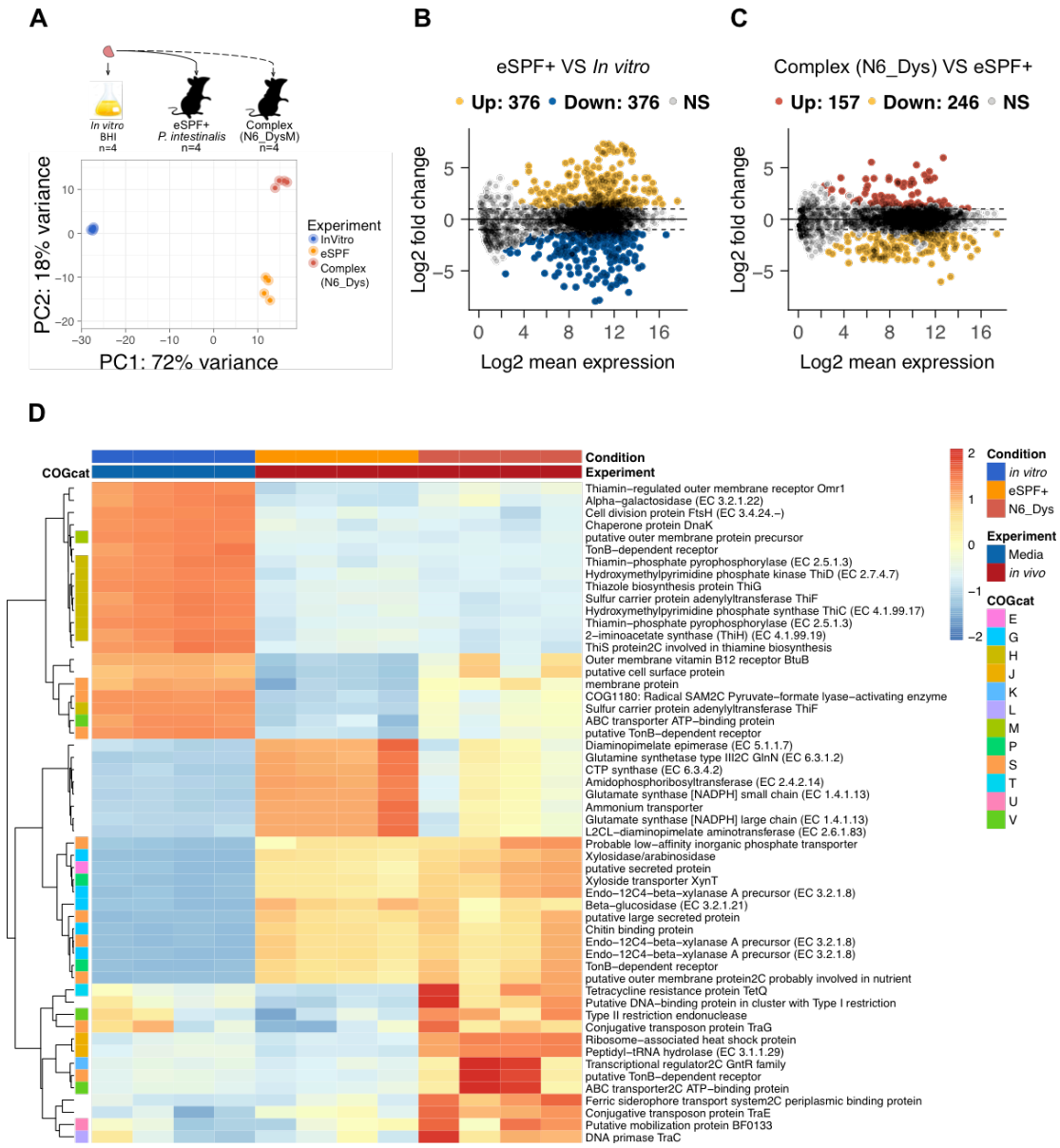

Figure 3.2.5: *P. intestinalis* transcriptome *in vitro* and *in vivo* after 3 weeks of colonization in two distinct microbiome models.

(A) Scheme of the treatments assessed by transcriptome analysis and principal component plot of gene expression profile by RNA-seq. (B, C) Log ratio and mean average plot (MA-plot,). (B) Gene expression levels of *P. intestinalis in vivo eSPF+ relative to in vitro* (BHI media). (C) Gene expression levels *in vivo* (complex N6_Dys) relative to eSPF+. Coloured dots indicate genes with log Fold Change > 2 and padj < 0.01. Yellow = Up-regulated in eSPF+, Red = Up-regulated in N6_DysM and Blue = Up-regulated *in vitro.* (D) Heat map of top 20 DEGs in each condition (log Fold Change > 5 and padj < 0.001), genes with unknown function were masked for heat map visualization. COG categories were identified for each gene: [M] Cell wall/membrane/envelope biogenesis, [U] Intracellular trafficking, secretion, and vesicular transport, [V] Defense mechanisms, [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination and repair, [E] Amino acid transport and metabolism, [G] Carbohydrate transport and metabolism, [H] Coenzyme transport and metabolism, [P] Inorganic ion transport and

(*Legend continued on next page*)

metabolism, [T] Signal transduction mechanisms, [S] Function unknown. Data include four experimental repeats. Iljazovic A, contributed with *Prevotella* colonization (Panel A).

3.2.6 The metabolic niche of *P. rodentium*

For *P. rodentium* I identified a total of 645 DEGs (397 up and 248 down) in eSPF+ mice vs *in vitro*. From the total set of DEGs, a portion of 36.3% corresponded to non-characterized CDS ("hypothetical protein" = 234, annotated = 411 CDS) (Figure 3.2.6B). The comparison of complex microbiota (Janvier) vs eSPF+ produced a total of 168 DEGs (108 up and 60 down) with a fraction of 39.2% of hypothetical CDS ("hypothetical protein" = 66, annotated = 102) (Figure 3.2.6C).

Analysis of overexpressed genes *in vitro* corresponded to cofactors and vitamins such as thiamine metabolism (*thiD* [EC: 2.71.49], *thiG* [EC:2.8.1.10], *thiH* [EC: 4.199.19]), DNA replication proteins (*ruvA* [EC:3.6.4.12], *dnaE* [EC:2.7.7.7], *hupA* DNA-binding protein HU) and transcription of key components in ribosome biogenesis and translation (L3, L7/L12, L9, L20, L28, L34, S6, S9, Peptidyl-tRNA hydrolase *pth* [EC:3.1.1.29]). The *in vitro* carbon metabolism was characterized by the up regulation of genes involved in galactose (*gla* [EC:3.2.1.22], *glk* [EC:2.7.1.2]) and fructose metabolism (*fruA* [EC:3.2.1.80]).

For the *in vivo* eSPF+ transcriptome, the genes associated to carbon metabolism were characterized by rhamnose (*rhaM*) and galactose metabolism via *lacZ* beta-galactosidase [EC:3.2.1.23]. Among the most expressed genes, I found an up regulation of genes involved in oxidative stress such as catalase-peroxidase [EC 1.11.1.6] and superoxide dismutase [Fe] (*sodB* [EC:1.15.1.1]). Interestingly, I found an up regulation of DMT-proteins (Drug/metabolite transporter), which are associated with

the transport of Queuosine and Archaeosine, modified nucleosides that are present in certain tRNAs in bacteria and archaea. The transcriptome in complex microbiome (Janvier) relative to eSPF+ was characterized by the up-regulation of several putative mobile proteins (BF0133) and hypothetical genes associated with conjugative transposons (*TraB, TraD, TraQ*) described previously in *Bacteroides* species (*133*) as well as stress elements such as the tetracycline resistant protein (TetQ).

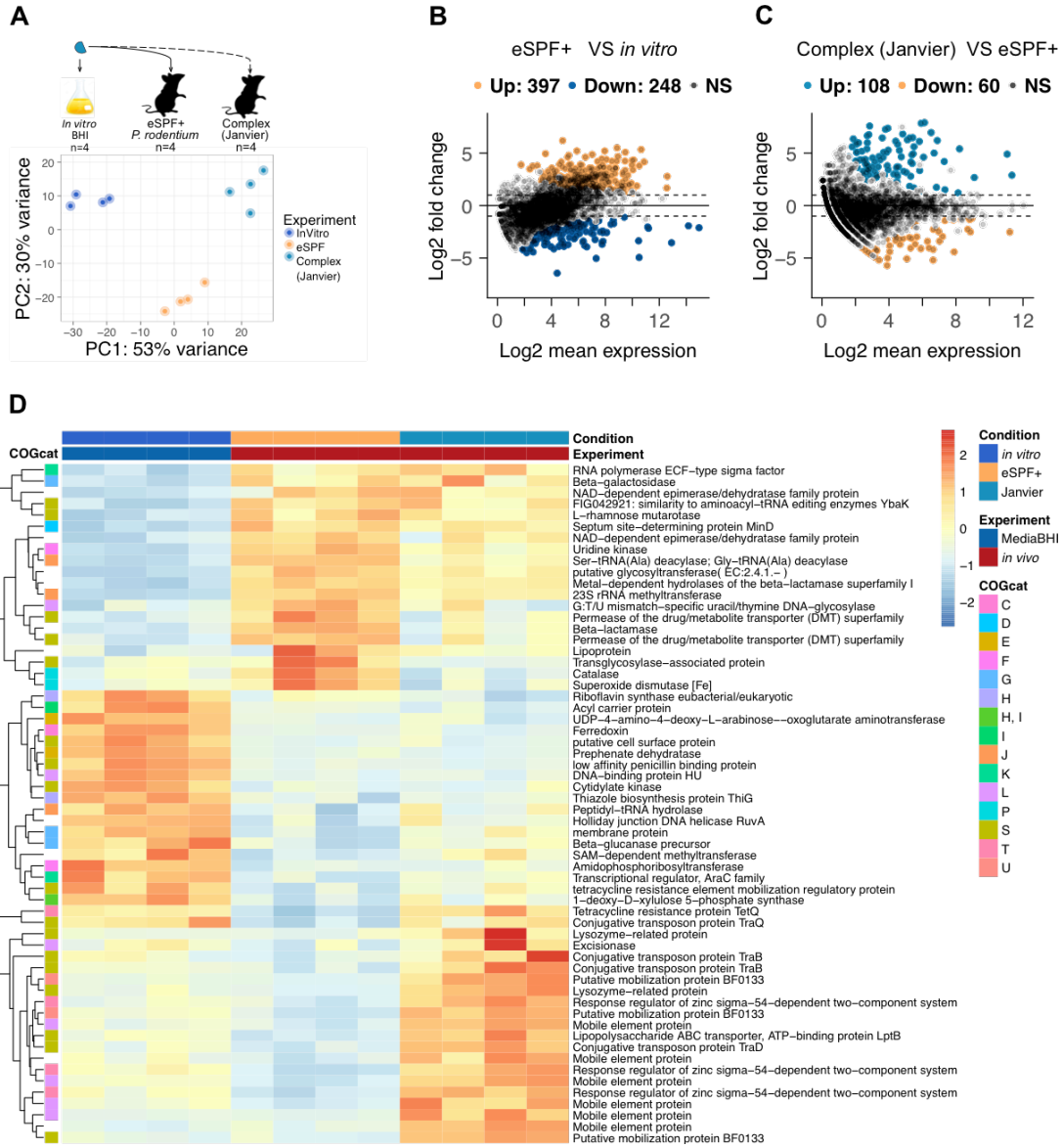

Figure 3.2.6: *P. rodentium* transcriptome *in vitro* and *in vivo* after 3 weeks of colonization

(*Legend continued on next page*)

in two distinct microbiome models.

(A) Scheme of the treatments assessed by transcriptome analysis and principal component plot of gene expression profile by RNA-seq. (B) Gene expression levels of *P. rodentium in vivo (eSPF+) relative to in vitro* (BHI media). (C) Gene expression levels *in vivo* (complex Janvier) relative to eSPF+. Coloured dots indicate genes with log Fold Change > 2 and padj < 0.01. Yellow = Up-regulated in eSPF+, Light blue = Up-regulated in Janvier and deep blue = Up-regulated *in vitro.* (D) Heat map of top 20 DEGs in each condition (log Fold Change > 5 and padj < 0.001), genes with unknown function were masked for heat map visualization. COG categories were identified for each gene: [D] Cell cycle control, cell division, chromosome partitioning, [T] Signal transduction mechanisms, [U] Intracellular trafficking, secretion, and vesicular transport, [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination and repair, [C] Energy production and conversion, [E] Amino acid transport and metabolism, [F] Nucleotide transport and metabolism, [G] Carbohydrate transport and metabolism, [H] Coenzyme transport and metabolism, [I] Lipid transport and metabolism, [P] Inorganic ion transport and metabolism, [S] Function unknown. Data include four experimental repeats. Iljazovic A, contributed with *Prevotella* colonization (Panel A).

3.2.7 The metabolic niche of *P. muris*

For *P. muris* I identified a total of 654 DEGs (382 up and 272 down) in eSPF+ mice relative to *in vitro* condition. From the total set of DEGs, a fraction of 42% corresponded to non-characterized CDS (“hypothetical protein” = 275, annotated = 379 CDS) (Figure 3.2.7). In complex microbiota (NCI1090) vs eSPF+, I identified a total of 144 DEGs (47 up and 97 down) with a fraction of 52.7% of hypothetical CDS (“hypothetical protein” = 76, annotated = 144) (Figure 3.2.7).

As observed for *P. intestinalis* and *P. rodentium*, the set of DEGs *in vitro* in *P. muris* corresponded essentially to genes involved in coenzyme transport and metabolism, specifically proteins associated with the thiamin biosynthesis (thiC [EC:4.1.99.17]; thiD [EC:2.7.4.71]; thiE [EC:2.5.1.3]; thiG [EC:2.8.1.10]; thiH[EC:4.1.99.19]), ribosomal proteins (S15, S20) and amino acid metabolism like the gene set for the conversion of L-histamine to L-glutamate (hutH EC:4.3.1.3]; hutU [EC:4.2.1.49]; hutI [EC:3.5.2.7]; fctD [EC:3.1.2.5]).

For eSPF+ I observed and up-regulation of genes encoding enzymes involved in ammonium assimilation such as *gdhA* glutamate dehydrogenase (NADP+) [EC:1.4.1.4] and glnN-2 glutamate synthetase type III-2 [EC:1.4.1.13, 1.4.1.14]. The carbon metabolism was characterize by the upregulation of genes involved in fucose and rhamnose catabolism such as the *fucU* (L−fucose mutarotase [EC:5.1.3.29]), L−fuco−beta−pyranose dehydrogenase (EC:1.1.1.122) and the rhamnulose-1-phosphate aldolase (*rhaD* [EC:4.1.2.19]).

Similar to *P. intestinalis* and *P. rodentium* in complex microbiome, the transcriptome of *P. muris* in complex (NCI1090) relative to eSPF+ was characterized by the up regulation of hypothetical mobilization genes (*bmgB*, *bmpH* BF0133) and putative genes associated with conjugative transposons (*traJ*, *traG*).

Together, these data unveil the metabolic responses of different *Prevotella* species under *in vitro* and *in vivo* conditions. In summary, we observed *in vitro* an up-regulation of genes involved in thiamine metabolism as well as the gene set involved in the central carbon metabolism (Embden-Meyerhof pathway). For *in vivo* eSPF+ we observed a different transcriptional landscaping where two main pathways were upregulated; the set of genes associated to the high affinity ammonia system, and a set of genes involved in the extracytoplasmic degradation of complex glycans such as xylan. In complex microbiome, the transcriptome in the three *Prevotella* species coincided in the up-regulation of stress response and mobile elements (transposon in Bacteroides). Collectively, these data support that species of *Prevotella* degrades complex polysaccharide and suggest that ammonia is utilized to assimilate nitrogen and synthetize amino acids.

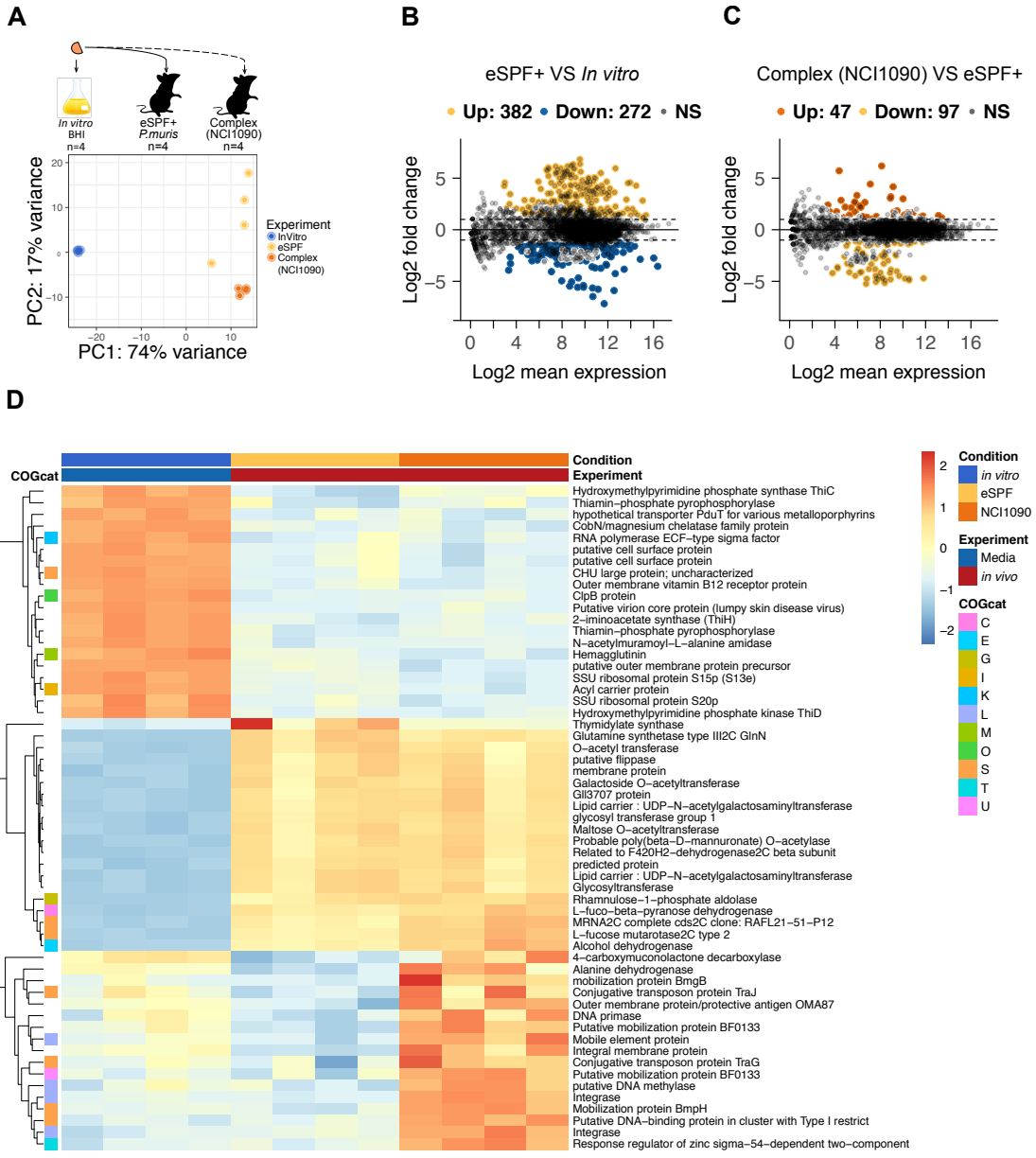

Figure 3.2.7: *P. muris* transcriptome *in vitro* and *in vivo* after 3 weeks of colonization in two distinct microbiome models.

(A) Scheme of the treatments assessed by transcriptome analysis and principal component plot of gene expression profile by RNA-seq. (B) Gene expression levels of *P. muris in vivo (eSPF+) relative to in vitro* (BHI media). (C) Gene expression levels *in vivo* (complex NCI1090) relative to eSPF+. Coloured dots indicate genes with log Fold Change > 2 and padj < 0.01. Yellow = Up-regulated in eSPF+, Orange = Up-regulated in complex (NCI1090) and deep blue = Up-regulated *in vitro.* (D) Heat map of top genes differentially expressed (log2 FoldChange > 5 and padj < 0.001). COG categories were identified for each gene: [C] Energy production and conversion, [E] Amino acid transport and metabolism, [G] Carbohydrate transport and metabolism, [K] Transcription, [L] Replication, recombination and repair, [M] Cell

(*Legend continued on next page*)

wall/membrane/envelope biogenesis, [I] Lipid transport and metabolism, [O] Post-translational modification, protein turnover, and chaperones, [S] Function unknown. [T] Signal transduction mechanisms, [U] Intracellular trafficking. Data include four experimental repeats. Iljazovic A, contributed with *Prevotella* colonization (Panel A).

3.2.8 Presence of distinct Glycoside Hydrolases (GHs) and expression of putative polysaccharide utilization loci *(PULs)* correlate with high fitness in *Prevotella* species

It has been proposed that dietary polysaccharides are one of the major factors in shaping the composition and physiology of the gut microbiome (*134*). In our previous transcriptional analysis, I identified a repertoire of carbohydrate degradation genes that were highly up-regulated *in vivo* by the isolated *Prevotella* species, specifically *P. intestinalis* presented a set of endoxylanases associated with xylan utilization. With the aim to investigate whether the carbohydrate degradation potential largely differs between *Prevotella* isolates and therefore could be associated with differences in fitness, I predicted and characterized the set of carbohydrate-active enzymes (CAZymes) for each *Prevotella* genome using two distinct bioinformatics approaches. First, I scanned the complete repertoire of CAZymes using the dbCAN2 meta server (*84*); this computational workflow integrates three state-of-the-art tools which significantly improve the annotation accuracy of the CAZyome (all CAZymes of a genome). Results of the CAZyome prediction unveiled that *P. intestinalis* presents the highest number of features accounting for a total of 186 CAZymes, followed by *P. rodentium* with 171 and *P. muris* with 163. From the total set of genes predicted, the most abundant enzyme classes in the *Prevotella* isolates were the carbohydrate-degrading genes GH (glycoside hydrolases) (Figure 3.2.10A). GHs are enzymes that catalyze the hydrolysis and/or rearrangement of glycosidic bonds. With the aim to

identify the differential GH degradation potential, I linked the GH families and its associated substrate for each strain. The results showed that GH43, GH2 and GH28 are the most abundant CAZymes in the *Prevotella* isolates. The GH43 is a complex group of endoxylanases, β-xylosidases and α-L-arabinofuranosidases which may play a key role in the de-branching and cleavage of xylans, while the GH2 family in bacteria has been commonly associated with β-galactosidases and β-glucuronidase activity. The GH28 family is a group of enzymes that are commonly active against rhamnogalacturonan, which forms the branched part of the pectin molecule.

Comparative analysis of GH families revealed that *P. intestinalis* contains a set of 7 GH families which are absent in *P. rodentium* and P. *muris*. The group of GH enzymes comprehends β-l-arabinofuranosidase (GH137), α-galacturonidase (GH138), α-2-O-methyl-l-fucosidase (GH139), α-l-fucosidase (GH141), DHA-hydrolase (GH143), α-L-rhamnosidase (GH145) and a complex of chitinases, lysozyme and xylanase inhibitors (GH18). Interestingly, GH137, GH138, GH139, GH141 and GH143 belong to the recently describe rhamnogalacturonan-II (RG-II) degradome, the most complex glycan currently known (Ndeh et al., 2017). Notably, *P. intestinalis* presented the highest amount of starch utilization system (*Sus* genes). The canonical *Sus* locus consists of eight adjacent genes, *sus*RABCDEFG, the *SusCD* gene pairs are considered as the functional core, which enables bacteria to bind and degrade specific carbohydrates. Using customized Hidden Markov Models (HMMs) designed for targeting conserved domains in *SusCD* homologs (*83*), I identified a total of 22 *SusCD*-like pairs in *P. intestinalis*, almost the double amount identified for *P. rodentium* with 13 *SusCD*, and *P. muris* with 8 pairs (Figure 3.2.10A).

The *SusC/SusD*-like pairs are common components of larger gene units known as polysaccharide utilization loci (PULs). Each PUL contains different enzyme clusters, which are tailored to recognize and degrade the extensive repertoire of complex carbohydrates presented in the mammalian gut. With the aim to reconstruct and identify the highly expressed PULs in our *Prevotella* isolates, I developed a bioinformatic tool based on above-mentioned HMMs to predict and quantify the expression of PULs based on genomic and transcriptome data. This program requires as input the gene location file (.gff), the amino acid sequences (.faa) and the gene expression data normalized in transcripts per million (TPMs). One important parameter "-length" needs to be defined; this parameter sets the maximum distance (in base pairs) between *SusC/SusD*-like pairs and the ultimate predicted CAZyme, allowing non-signature genes to occur in between the PULs. After the genes are grouped into putative operons, the average PUL transcription is calculated and transformed to log2-fold-change with respect to the median of the whole genome transcriptome. Our computational tool reconstructed 14 PULs for *P. intestinalis*, in line with our previous CAZyome characterization; this isolate presented the highest amount of PULs in comparison with *P. rodentium* that contained 10 PULs and *P. muris* with 6 PULs (Figure 3.2.10C). Using the transcriptome data, I ranked and identified the most expressed PULs in vivo for each of the *Prevotella* isolates. Notably, I found that the most expressed PULs in *P. intestinalis* (PUL1, PUL2, PUL3) are absent in *P. rodentium* and P. *muris*. Based on the grouped trophic guilds for each CAZymes (*42*), the predicted PULs are associated with the degradation of plant-based polysaccharides, specifically, the PUL2 that contains a tandem repeats of *SusC/SusD* homologues has been previously characterized as an essential locus for xylan catabolism. Phylogenetic analysis based on the amino acid

sequences of all *SusC/SusD* confirmed that the *SusCD* genes from the most expressed PULs in *P. intestinalis* clustered separately to the most expressed PULs in *P. rodentium* and *P. muris* and therefore are specific elements present in the highly competitive *P. intestinalis* isolate (Figure 3.2.10D-E).

Our reconstruction of the carbohydrate catabolism in *Prevotella* suggested that *P. intestinalis* encodes and expresses divergent Xylan- and Pectin-degrading genes attributable to the versatility of the CAZymes and SusC/D homologs encoded in PULs and this functional difference could provide a competitive advantage to *P. intestinalis in vivo* in the presence of complex carbohydrates like xylan and pectin.

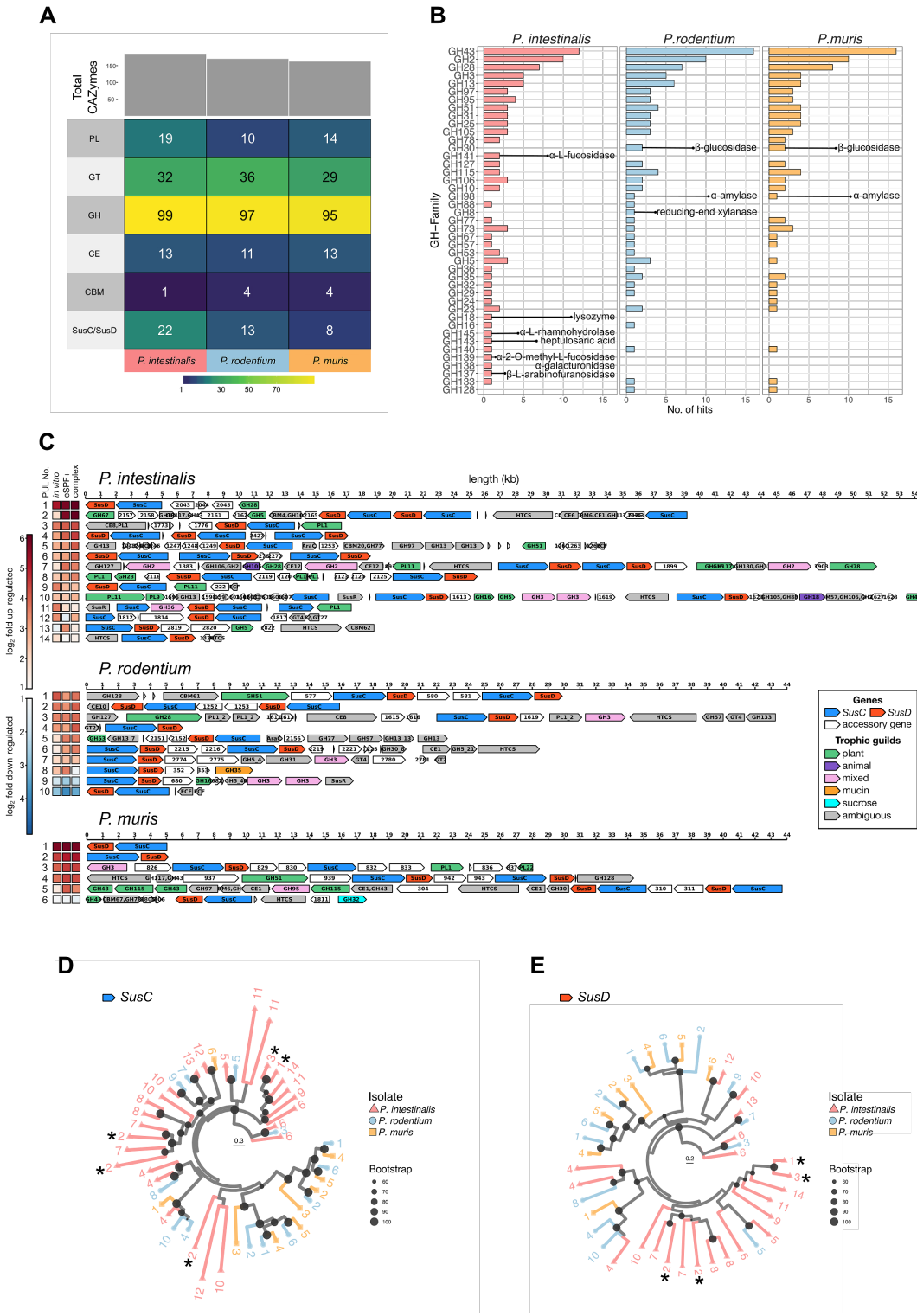

Figure 3.2.8: CAZymes repertoire and expression of Polysaccharide Utilization Loci (PULs) in *Prevotella spp*.

(A) Heat map indicating the number of CAZymes and SusC/SusD pairs in *Prevotella* genomes. Enzyme classes are listed as Polysaccharide Lyases (PLs), Glycosyl Transferases (GTs), Glycoside Hydrolases (GHs), Carbohydrate Esterases (CEs), Carbohydrate-Binding Modules (CBMs) and SusC/SusD pairs. (B)

(*Legend continued on next page*)

Glycoside Hydrolases (GH families) distribution in *Prevotella* spp. Associated functions of differentially presented GH enzymes were highlighted. (C) Prediction and expression of PULs in each *Prevotella* genome. Colors highlight the location of SusC/SusD gene pairs (Blue and Orange) and the CAZymes potential substrate. (D) Phylogenetic tree analysis based on *SusC* amino acids sequences. (E) Phylogenetic analysis of *SusD* proteins. Tree was constructed by the RaxML algorithm with 1000 bootstrap replicates. The branches were colored by species and tip labels represent the PUL group where the gene was predicted. Stars highlight the SusC/SusD genes associated to the top 3 most expressed PULs in *P. rodentium.*

3.2.9 Diet rich in Xylan and Pectin modulates the stability of *Prevotella intestinalis* in the mouse gut

I identified that putative PULs correlate with the high fitness in *Prevotella* species, especially those associated with xylan degradation. With the aim to investigate whether complex carbohydrates or microbial interactions are required for *P. intestinalis* colonization and stability we used two distinct approaches. First, we performed an *in vivo* competition experiment inoculating the same mixture of strains (*Prevotella intestinalis* and *Prevotella rodentium*) into WT mice harboring the eSPF microbiota. Mice were kept with standard chow rich in complex plant polysaccharides (stand-PP) before and after the animals received a single oral gavage of the strain's mixture. Furthermore, eSPF+*Prevotella* mice were divided in three different treatments (n=4), a control maintained with stand-PP and two experimental groups where the diet was switched during 7 days to one of the two semisynthetic diets rich in high fat (Synth-HF) or a control rich in simple sugars and low-fat (Synth-LF). After 1 week of diet intervention, both experimental groups were returned to regular chow (stand-PP) (Figure 3.2.10A and D). Second, using the same diet intervention set-up, we assessed the influence of complex microbial communities on the stability of *Prevotella* strains by mixing in equal proportions the total gut microbiome where *P. intestinalis* and *P.*

rodentium were originally found in high abundance. Then we used 16S rRNA gene sequencing from fecal pellets collected over the course of each experimental group to identify the relative abundance of *P. intestinalis* and *P. rodentium* as well as the changes in the total microbial community.

Analysis of β-diversity using principal coordinate analysis (PCoA) showed that colonization with *Prevotella* isolates reshapes the microbial community structure in eSPF mice after 2 weeks of colonization (Figure 3.2.10A, Time point D-1). Following *Prevotella* colonization, the control group remained stable during the time course, while the experimental group exhibited consistent diet-induced changes in the microbiota during the time points of diet intervention. Interestingly, switching the mice between Stand-PP/Synth-HF and Stand-PP/Synth-LF rapidly alters the abundance of *Prevotella*ceae (Figure 3.2.10B, D1, D3 and D7). A detailed analysis of the relative abundance of the *Prevotella* isolates showed an initial dominance of *P. intestinalis* over *P. rodentium* during diet oscillations (Figure 3.2.10C). For *P. intestinalis* a diet switch from Stand-PP into Synth-HF or Synth-LF resulted in a severe reduction in relative abundance (45.10%, SEM=5.38 to 3.35%, SEM=1.04 and 29.97%, SEM=3.65 to 7.61%, SEM= 2.54 respectively). The reduction of *P. intestinalis* did not lead to an increase of *P. rodentium* and its relative abundance also decreased during Synth-HF and Synth-LF intervention (11.10%, SEM=2.35 to 0.77%, SEM=0.08 and 9.43%, SEM=2.18 to 1.39%, SEM= 0.16 respectively). Notably, returning the mice to Stand-PP diet rapidly increased the relative abundance of *Prevotella* isolates after one day of recovery; particularly *P. intestinalis* recovered its predominant abundance from 1.00% to 27.39% (SEM = 4.88) in Synth-HF and from 3.89% to 48.90% (SEM 4.45) in Synth-LF. Analysis of the microbiome composition at the level of family unveiled distinct response

patterns, for the initial colonization experiment using the mixture of *P. intestinalis* and *P. rodentium* isolates; a reduction of Rikenellaceae and Lachnospiraceae was observed upon colonization. Interestingly, during the diet intervention Ruminococcaceae and Porphyromonadaceae were expanded in Synth-HF and Synth-LF correspondingly (Figure 3.2.10B).

In the microbiome experiment, I observed similar dynamics in the assembly of the *Prevotella* species. In contrast to the isolate colonization, the microbiome experiment co-transferred a complex mixture of Bacteroidales (S24-7, Bacteroidaceae) and Proteobacteria (*Desulfovibrio* and *Helicobacter*) (Figure 3.2.10F). In line with our previous observation (see Figure 3.2.4), *P. intestinalis* dominates in the gut of mice fed with standard chow diet even in the presence of complex microbial communities (Figure 3.2.10E). In comparison with the experiment using the *Prevotella* isolates in eSPF mice, diet intervention in mice colonized with the complex microbiome showed a decrease in *Prevotella spp.,* and an expansion of Bacteroidales (S24-7 and Bacteroidaceae) as well as Ruminococcaceae in Synth-HF and Synth-LF. Notably, I observed a delayed recovery of *P. intestinalis* after returning the mice to stand-PP in Synth-HF and in Synth-LF compared with the recovery in the mice colonized with the mixture of single isolates.

Together, these experiments suggest that complex plant polysaccharides are utilized by *Prevotella* strains and they serve as an essential substrate to maintain interspecies stability. We conclude that complex gut microbiota rich in Bacteroidaceae species and diet base in simple sugars strongly reduce the abundance and fitness of *P. intestinalis*.

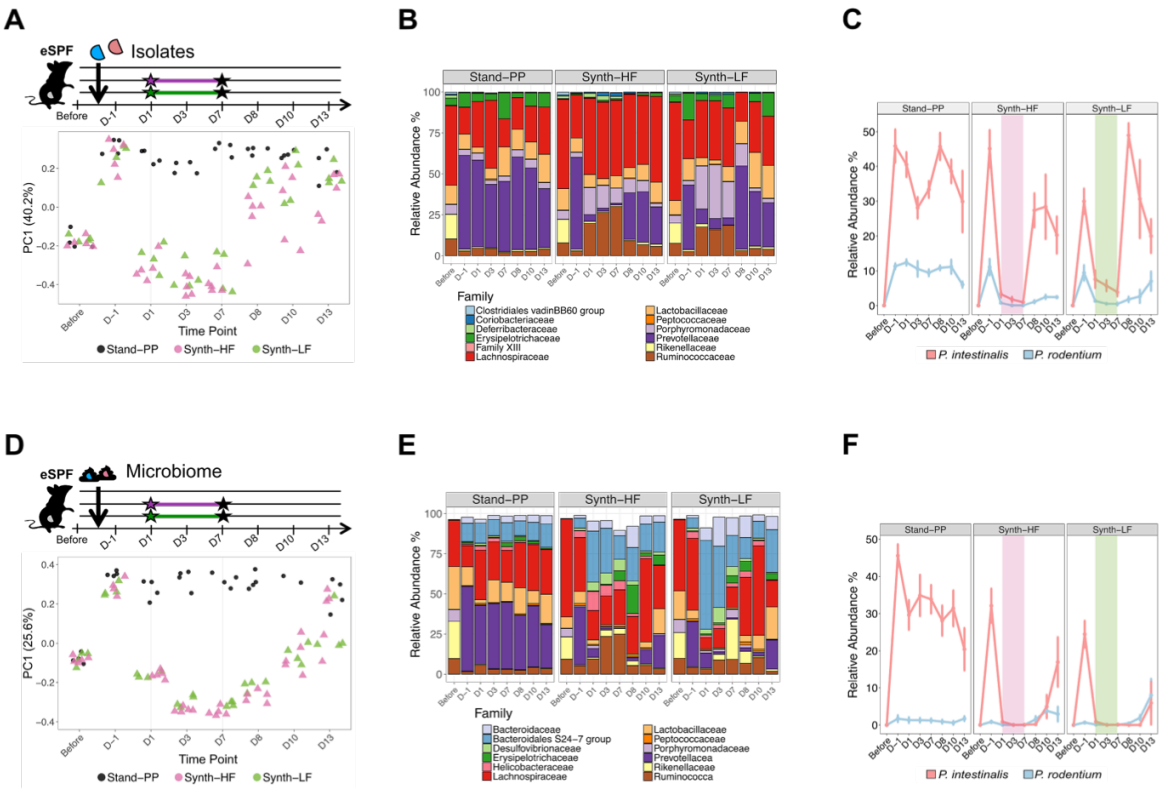

Figure 3.2.9: The effect of diet intervention and microbial communities on the stability of *P. intestinalis.*

Diet intervention experiment in eSPF mice using an equal mix of *Prevotella* strains. (A) Experimental design and similarity of gut microbial communities using PCoA ordination analysis. Each shape represents a fecal sample from a mouse belonging to the indicated diet group; y-axis indicates PCoA1 using Bray-Curtis distances and x-axis indicates the time point at which diet switch occurred. (B) Relative abundance of bacterial Families for each diet intervention group collected at the indicated time points. (C) The relative abundance dynamics of *P. intestinalis* (in red) and *P. rodentium* (in blue). Diet intervention experiment using a microbiome mix where the *Prevotella* strains where isolated. (D) Experimental design and PCoA ordination analysis. (E) Relative abundance of bacterial Families in complex microbiome. (F) Relative abundance dynamics of *P. intestinalis* (in red) and *P. rodentium* (in blue). Iljazovic A, contributed with *Prevotella* colonization and FT (Panel A and D).

4 General discussion and outlook

The overall objective of this research was to investigate the interplay host and the microbiota focusing on two aspects: first, the contribution of the host immune system on shaping the gut microbiota and second, to characterize an important bacterial family present in the intestinal microbiota, the *Prevotella*ceae, which has been associated with health and disease. In the following I will discuss the obtained insights in the context of recent observations in this dynamically evolving research field.

The gut microbiome is a complex environment. It has been described as one of the most densely populated habitats with an estimate of 0.15 kg of bacterial biomass (*4*). As a result of the recent improvements in the next generation sequencing (NGS), the characterization of the structure “who is there” and their potential functionality “what are they doing” are now pieces of information that can be inferred with precision and high throughput. However, despite the exponential increase of publications from 2010 with 380 articles per year to 4249 in 2017 (PubMed search with the terms “gut microbiota” in 2018), the identification of individual factors contributing to the composition of “healthy” and “dysbiotic” microbiome remain relatively understudied. Currently, one of the biggest challenges in microbiome research is the translation of association studies into causative relationships through a combination of functional “omics” and mechanistic studies of host responses in animal models or interventional clinical trials. During the experimental studies of the here presented thesis, I have used murine models with a standardized microbiota in combination with functional studies to collect results and evidence to move beyond conventional association into a functional understanding.

4.1 Enhanced specific pathogen free mice (eSPF)

One of the key aspects in experimental sciences to enhance reproducibility is the standardization of experimental conditions, which poses specific challenges with regard to standardizing animal models. As an example of the vast differences between SPF mice, a recent microbiome survey found a large variation in the C57BL/6J fecal microbiota across different German animal facilities. The findings of that study suggest that microbiome variation is influenced by factors such as the use of irradiated chow, the open cage system and husbandry practices in each animal facility (*102*).

Hence, controlling or standardizing the microbiome in animals is a very challenging task, first, because it is frequently necessary to perform generational experiments in distinct genetic backgrounds to characterize the stability and dynamics of the microbial communities and, second, because the murine gut microbiome remains largely less characterized than the human microbiome. Despite the challenges in keeping stabilized bacterial communities in laboratory animals, standardization of the microbiome and the use of littermate-breeding controls are some of the reliable approaches to distinguish the relative effect of host-genetics and gut microbiome in the contribution of phenotypic traits. (*135*). In addition, the isolation and functional characterization of prevalent commensals in combination with standardized animal models could allow researchers to dissect phenotypes from correlation to causality.

In the past years, different experimental models have been developed with the aim to determine if a specific phenotype is mainly driven by the microbiome or host genetics. Currently, the best-known and most widely used animal models are gnotobiotic and isobiotic animals. One of the earliest gnotobiotic animal model is the low-complexity

ASF community (*136*, *137*), but recently new models have been established such as the Oligo-Mouse-Microbiota (OMM12 mice) (*138*). While studies with classical gnotobiotic animals have provided important insights about microbiome functionality (*139*), it might be important to consider that these studies not always resemble the physiology of animals with a complex microbial community, since the ASF mice display an incomplete metabolic functionality to what is found in SPF animals harboring a complex microbiome (*140*). Hence, in order to overcome the limitations of low-complexity gnotobiotic models, researchers recently proposed a compendium of standardized criteria to establish animal complex microbiotas and generate reproducible isobiotic models (*2*). In brief, the criteria were defined in 9 different principles listed as follow: i) microbial stability over generations, ii) fully maternal transferability, iii) reconstitution in germ-free wild-type mice, iv) completeness of microbial metabolic pathway inferred from the total metagenome, v) broad taxa representation, vi) no physiological abnormalities in wild-type mice, vii) similar metabolomic and immunological profiles with mice harboring a diverse microbiota, viii) normal pathogen colonization resistance and inflammatory model disease susceptibility and ix) relative stability of the isobiotic microbiota under aseptic husbandry within individually ventilated cages.

The results presented in section 3.1 together with our previous publications confirmed that eSPF-HZI microbiota featured most of the previous criteria as a standardized animal model (see criteria I, ii, iii, v, vi, vii, viii and ix) (*2*). In addition of the microbial stability in distinct immune-deficient mice and over generations (*63*, *141*), we recently characterized a non-pathological role during induced inflammation (*142*) and a normal pathogen colonization resistance in comparison with other non-standardized SPF models (*143*). Altogether this highlights that eSPF mice represent a stable and

reproducible *in vivo* model to study gut microbiota interactions allowing researchers to investigate the influence of key bacterial members on host physiology as well the influence of the host-genetics on the control of pathobionts. Ongoing studies are focused on the isolation and characterization of the microbial metabolic potential using targeted-anaerobic culturing and deep sequencing metagenomics (see criteria iv).

4.2 The effect of Nlrp6 and adaptive immunity on microbiota composition depends on community structure

We used the eSPF model to assess the effect of the NLRP6 sensor on shaping the gut microbiota. Previously, significant fecal microbiota alterations were observed in NLRP6-deficient mice raising the possibility that deficiencies in the Nlrp6 inflammasome pathway, i.e., innate immunity, may underlie dysbiosis-associated disorders and therefore exacerbate colitis (*47*, *106*, *144*). In contrast, a recent study by Mamantopoulos and collaborators (*95*) and another by Lemire and colleagues (*98*) analyzed the microbiome of Nlrp6-deficient mice using SPF littermates that were separated after birth. The authors did not find any taxonomical difference in microbial composition or colitis susceptibility, concluding that previous genotype-dependent effects on microbiome composition were driven by maternal transmission and cage covariates.

In contrast, we hypothesized that these conflicting results could be explained by differences in the microbiome composition and therefore the contradictory results may reflect the absence of dysbiotic-triggering microbes in a given experimental

environment. To test this hypothesis, we first surveyed the microbiome composition of several colonies of *Nlrp6*$^{-/-}$ mice raised in a conventional animal facility, our results support that beyond host-genetics, additional factors such as maternal inheritance and stochastic events shape the gut microbiome composition in conventionally housed mice, even when they are exposed to a similar environment within the same animal facility. Next, the aforementioned *Nlrp6*$^{-/-}$ and immunodeficient *Rag2-/-* mice were rederived into eSPF conditions by embryo transfer. Our results showed that under eSPF conditions, *Nlrp6*-dependent effects on the microbiome were not detectable, which agrees with the work by Mamantopoulos and colleagues. In addition, our experiments on eSPF mice highlighted that even mice deficient in adaptive immunity (*Rag2-/-*), which have been described to feature microbiome aberrations (*94*, *99*, *100*) presented no differences in microbial communities under eSPF conditions. These data together suggest that a gut microbiota depleted of pathobionts can be maintained stable even in immunocompromised mice within a facility with well-trained staff and following strict handling procedures.

Subsequently, we aimed to test whether exposure to pathobionts may cause imbalances in the microbiota of *Nlrp6*$^{-/-}$ mice. To answer this question, we performed a microbiome challenge by fecal transplant (FT) introducing a dysbiotic community from conventionally housed *Nlrp6*$^{-/-}$ mice (N6_Dys) into isobiotic WT, *Rag2*$^{-/-}$, and *Nlrp6*$^{-/-}$ mice harboring the same eSPF microbiota. Our detailed characterization of the microbiome from samples in distinct intestinal locations confirmed a significant contribution of the genotype to microbiome composition. For the *Nlrp6*$^{-/-}$ mice, multivariate analysis (ADONIS) accounted that 10% of the microbiome variability in the proximal colon (PC) and 13% in the distal colon (DC) could be explained by a deficiency

in the Nlrp6 inflammasome. For the *Rag2*$^{-/-}$ mice, the observed effect was larger. In the PC the microbiome variability was explained by 27% and in the DC by 32%. These results are in line with the nature and immunodeficient degree of the Rag2 model, which lacks both B and T cells and consequently several components that have been previously characterized in shaping the microbiome such as IgA (*104*). In the case of the Nlrp6 sensor recently reviewed by Levy *et al.,* 2017 (*145*), it has been associated with the regulation of goblet cells, the secretion of antimicrobial peptides and the recognition of viral infections, however the molecular mechanism of the triggers remain unknown.

Taking our results together, our conclusion differs from that of Mamantopoulos *et al.,* and Lemire *et al.,* who reported that “inflammasomes do not affect the gut microbiota composition”. Instead, our experiments shed light in the understanding of the inflammasome activation and resemble previous observations in human cohorts, where chronic inflammatory conditions are the outcome of a multifactorial trait between host genetics, specific microbes and environmental factors (Figure 4.1). Finally, we believe that our study reconciles the opposing results that have been reported regarding the role of Nlrp6 and, potentially, other immune components in shaping the microbiome. As a conclusion from our experiments, it is strongly recommended to evaluate mice phenotypes of host-microbiome interactions by using standardized microbiota as well as complex microbiome with the aim to estimate and understand the contribution of genetics in eubiosis and dysbiosis. Moreover, it is important to assume that WT and KO mice that have been bred separately in a conventional animal facility do not share the same microbiota and, therefore, microbiome analysis is recommended to estimate variability. The best alternatives for microbiome normalization are: i) rederivation by

embryo transfer, ii) littermate controls or iii) microbiota synchronization by co-housing for at least 4-6 weeks (*146*).

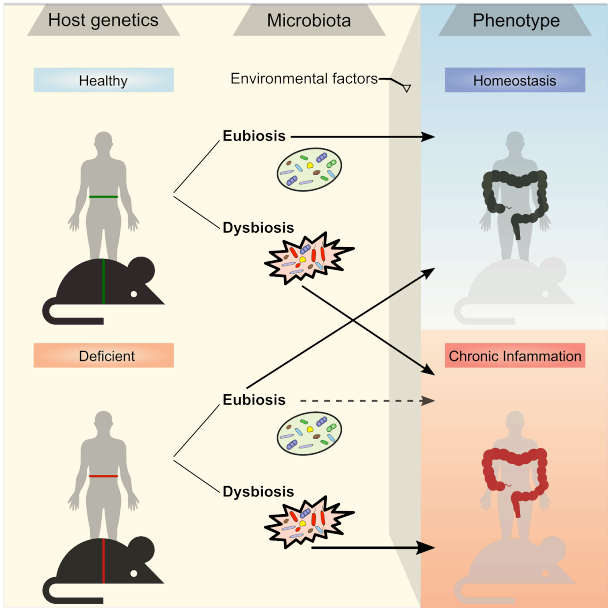

Figure 4.1: Model for host-genetics and gut microbiota interactions in the developing of disease.

4.3 Environmental factors dominate over host genetics in shaping the gut microbiome

Besides the impact of host-immune deficiencies in the gut microbiota, Ruth Ley and colleagues (*49*, *50*) documented in a twin cohort that genetic background in humans strongly correlates with the abundance of certain health-associated bacteria. Whether these correlations are a consequence of genetics or environmental factors remains elusive. In contrast, recent analyses on large human cohorts have observed

that environment dominates over host genetics in shaping the gut microbiome (*147*). Especially diet (*53*) and host-geographical location (*58*) are among the major environmental drivers of gut microbiota variability. Specifically, Turnbough and collegues (*54*) previously determined in animal models that diet dominates over genetics in shaping the murine gut microbiota. Using multivariate models (ADONIS test) the authors described that diet (High-Fat/High-Sugar vs Low-Fat/Plant-Polysaccharide) explained a fraction between 35% to 48% of the observed variation, while deficiencies in innate- (*NOD2*-) and adaptive- (*MyD88*- and *Rag1*-) immunity explained a lower fraction ranging from 8% to 23%. This previous report agrees with our findings where the influence of host-immune deficiencies could contribute from 13% for innate- (NLRP6) to 30% for adaptive- (Rag2) immunity.

4.4 Diversity and functional niche of prevalent species of intestinal *Prevotella* isolated from mice.

Initial microbiome surveys in human cohorts already described that diet strongly modulates the configuration of the gut microbiome, particularly protein and animal fat were associated with a predominance of *Bacteroides* and plant carbohydrates with *Prevotella* (*53*, *148*). Furthermore, distinct studies have found an association between *Prevotella* prevalent microbiome in rural populations who consume a plant-rich diet (*149–151*). The observations of a high prevalence of *Prevotella* in non-westerns (populations linked with vegetarianism) have suggested that *Prevotella* is a beneficial gut commensal. However, an overexpansion of *Prevotella* in the gut has also been

linked with autoimmune diseases (*121*) and exacerbation of experimental colitis in mice (*152*). One of the hypotheses to explain the contradictory phenotypes of *Prevotella* associated with health and disease is the large genome diversity of the strains within the genus and between *Prevotella* species (*123*). Yet, the genome and functional characterization of reference strains and the genetic basis of the prevalence of *Prevotella* species in the gut remain to be understood. The section 3.2 in this thesis aimed to characterize the diversity and functional role of four novel *Prevotella* spp. isolated from the mouse gut. The isolated species were named as follow: *Prevotella rodentium sp. nov., Prevotella muris sp. nov., Prevotella intestinalis sp. nov.*, and *Prevotella musculis sp. nov*., based on their abundance and prevalence in different hosts according to the sequence read archive SRA NCBI database.

The physiological implications and potential immunomodulatory effects on the host by species of the family *Prevotella*ceae remain elusive, yet somewhat opposing results have been reported regarding the impact of intestinal *Prevotella* species on the host. In a prebiotic intervention trial, Kovatcheva-Datchary and colleagues reported an improvement of glucose metabolisms in individuals with *Prevotella* positive microbiome, while no improvement was observed in individuals without *Prevotella* colonization suggesting that *Prevotella* provides health beneficial effects. Yet, in other studies the presence or relative overabundance of different intestinal *Prevotella* phylotypes has been linked with chronic inflammatory diseases (*46*, *47*). Currently, the differences observed in responses to diet and health conditions across studies have been hypothesized to be explained by the potential high diversity within and between *Prevotella* spp. However, genome-wide analysis on distinct cultivable-reference species

and the identification of the diversity in the accessory genome, have yet to be described.

In agreement with similar metagenomics studies in humans cohorts (*22*), our comparative genome analysis in combination with 16*S* rRNA gene diversity and culture-based approaches, demonstrated the vast diversity presented in the genus *Prevotella*. Despite the large diversity observed, our results also demonstrated a strong inter-species antagonism between the species of *Prevotella* and members of *Bacteroides*. This competition has been previously proposed where members of the *Prevotella* or Bacteroidetes are present but rarely a coexistence of both strains (*153*). Using controlled *Prevotella* competition experiments in mice with standardized microbiota (eSPF), we confirmed a high interspecies competition. In addition, we observed that a specific specie of *Prevotella* is able to overtake and influence the proportion of all *Bacteroidetes* in the gut (*154*, *155*). In order to identify the metabolic niche of our newly isolated *Prevotella* species collection and the potential features associated with dominance, we characterized the transcriptional landscaping of three distinct *Prevotella* species *in vitro* and *in vivo* conditions.

Our RNA-seq analysis of the three isolated intestinal *Prevotella* species displayed both shared and distinct changes in their transcriptional landscape when they were exposed to three different conditions: i) *in vitro* BHI+ media, ii) *in vivo* with mice with a standardized microbiome eSPF and iii) *in vivo* in mice with different complex microbiomes. The shared set of most up-regulated genes *in vitro* corresponded to the operon for thiamine biosynthesis. It has been shown that thiamine is necessary for glycolysis, the tricarboxylic acid (TCA) cycle, branched-chain amino acid metabolism, and nucleotide metabolism (*156*). Specifically, thiamine is required in the (TCA) cycle

for the synthesis of acetolactate and this reaction is performed by the Acetohydroxyacid synthase, which utilizes thiamine pyrophosphate in order to link two pyruvate molecules. The up-regulation of this pathway *in vitro* by *Prevotella* spp., could suggest a low level of thiamine in the BHI+S media and a supplementation with thiamine could potentially improve the isolation and growth of unculturable *Prevotella*ceae species.

Our experiment *in vivo* under eSPF microbiota conditions revealed a high up-regulation of glycoside hydrolases, specially endo-1,4-beta-xylanases [EC:3.2.1.8] belonging to GH43 CAZymes and several genes encoding proteins with unassigned functions. A similar functional profile has been found in a previous study in rumen *Bacteroides* (*118*). In addition, we observed a high expression of genes encoding for the high-substrate affinity nitrogen assimilation enzymes GS-GOGAT pathway (*157*), this is in line with previous observation of rumen bacteria where ammonium is the preferred source for growth (*158*). Previous work in *Prevotella ruminicola 23* highlighted that expression of GS-GOGAT system could be induced *in vitro* under growth-limiting and non-limiting nitrogen concentrations (*159*). It is also known that undigested starch reduces fecal concentrations of ammonia and phenols in humans (*160*). Together, this may reflect an adaptation of *Prevotella* species to low levels of ammonia in the gut environment. Moreover, the expression of the ammonia high affinity system together with the set of genes for the degradation of complex plant polysaccharides has been also observed in sulfate-reducing bacteria such as *Desulfovibrio piger* (*161*), indicating that the genes encoding for the degradation of complex carbohydrates and nitrogen metabolism are co-expressed in distinct bacteria phyla and are important gene sets for bacterial fitness in the gut.

The differential gene expression of *Prevotella* species in a complex microbiome environment compared to a community with potentially fewer competitors unveiled the up-regulation of genes associated with stress elements, transposons and mobile elements. Recently, it was shown that the gut mucus layer provides a distinct microbial niche and therefore shaped the transcription profile in commensal bacteria such as *B. thetaiotaomicron* and *E. coli* (*162*). But the effect of differences in microbial composition on the transcriptome of particular gut commensals has been less well studied. In our results we did not find evidence of an alternative carbohydrate preference depending on the community structure, however, the three species of *Prevotella* coincided with an up-regulation of conjugative transposons, whose activation is associated with tetracycline resistance (*163*). This could indicate a potential sensing mechanism between Bacteroidetes, which activates the expression of self-transmissible elements or transposon inhibition by antisense RNAs (*164*). The study of mobile elements in *Prevotella* has not been explored yet, and the understanding of the molecular bases could lead to the design and improvement of molecular tools for functional genomics.

The transcriptome analysis of *Prevotella* species *in vivo* also highlighted the expression of several genes clusters with unassigned function occurring next to glycoside hydrolases. Therefore, we developed a bioinformatics tool to identify, reconstruct and quantify the expression of PULs using genomic and transcriptome data. Our results unveiled that the total number of CAZymes and SusC/SusD homologous genes highly differ between the three *Prevotella* isolates, and the expression of putative xylan-degrading PULs could represent a genetic determinant for enhanced fitness in *P. intestinalis*. Previous studies have started to unveil PUL functioning and substrate specificity in Bacteroides (*10*, *83*, *134*, *165*). The PUL_02 in *P. intestinalis* resembles a

PUL up-regulated in *Bacteroides ovatus ATCC 8483T* when it was grown on xylan, the so called PUL-XylL (*36*, *61*). The over-expressed PUL_01 and PUL_3 in *P. intestinalis* did not present a high similarity with previous characterized PULs and the closest homologue found was the PUL 59 in *Bacteroides vulgatus ATCC 8482* (BVU_2919 - BVU_2922). Base on the CAZymes gene annotation, the PUL_01 contains a gene that encodes for a GH23 enzyme, which is involved in the degradation of polygalacturonases and rhamnogalacturonan. Interestingly, the PUL_03 that contains homologues SusC/D genes with PUL_01 presented a Polysaccharide Lyases gene (PL1) that is associated with the degradation of pectate and pectin. These results together support previous association between a *Prevotella*-dominated microbiota and complex plant polysaccharides intake (*44*, *166*).

Apart from cellulose and hemicellulose, pectin utilization by *Bacteroides* and *Prevotella* has been less investigated; recently a previously uncharacterized *Prevotella* species (*P. pectonivorans)* was studied by the ability to grow on pectin as a single carbon source. Using genome sequencing a group of CAZy families (CE8, CE12, GH28, GH78, GH105, PL1, PL22) were predicted as potential gene units responsible for the pectin degradation (*167*). Our results from the prediction of the CAZyome showed that *P. intestinalis,* besides the GH28 and PL1, encodes an additional set of recently described enzymes which are able to metabolize rhamnogalacturonan-II (RG-II) independently of *susC/D* genes (*168*). Together, these results indicate that *P. intestinalis* encodes specialized gene machinery for the degradation of hemicellulose, specifically xylans and pectins (RG-II) (Figure 4.2).

In summary, our results highlight the vast genomics diversity in the genus *Prevotella* and uncover the limitations of functional predictions based on 16S rRNA gene.

Integration of shallow metagenomics in combination with longitudinal diversity experiments and culture-based microbiology has been proved to be a powerful approach to understand the effect of carbohydrates in the stability and resilience (ability of a system to absorb changes and still persist) of dominant gut commensals. Given the significant contribution of the PUL diversity in determining *Prevotella* fitness and species dynamics, dietary interventions with specific substrates such as arabinoxylans and RG could have a considerable potential as a method for modulating or establishing engineered microbial species in the gut, when therapies like fecal transplantation against *Clostridium difficile* infection fails due to the instability of the transplanted microbial antagonists.

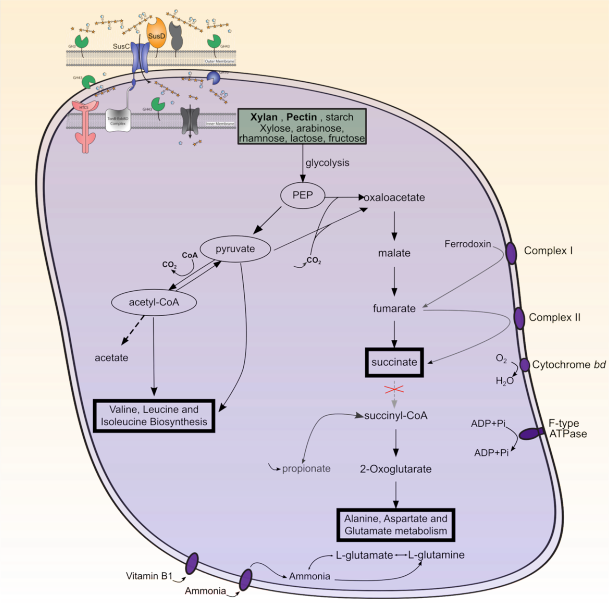

Figure 4.2: Model of key metabolic pathways of intestinal *Prevotella* spp. inferred from in vivo transcriptome.

Circles indicate intermediate products and squares indicate putative final products. Arrows indicates the direction of the reactions.

4.5 General Conclusion and Perspectives

In conclusion, I have utilized 16*S* rRNA diversity analysis in combination with genomics and functional assessments as powerful tools to close the gap between correlations and causality. In the first part of this thesis I have determined the influence of host genetics in shaping the microbiome using a standardized microbiota model eSPF. Although, my results help to understand the contribution degree of NLRP6 in shaping the gut microbiota, several questions outstand from our results. What are the specific ligands (metabolism or antigens) that activate the NLRP6? How do certain species of *Helicobacter* explore a specific immune-niche in the NLRP6-deficient mice and what is the long-term effect of NLRP6 deficiency on the gut microbiota composition in a generational study?

In the second part of this thesis, I have combined multi-omics approaches to characterize the diversity of *Prevotella*ceae in mice and understand the metabolic niche and fitness of novel isolates of *Prevotella*. I unveiled that specific polysaccharide utilization loci PULs in combination with a diet rich in complex carbohydrates are strongly linked with the predominance of *Prevotella* spp in the mouse gut. Future experiment will aim to engineer *Prevotella* strains with the aim to identify the minimal gene systems responsible of high fitness in the gut. Finally, our understanding of the interplay between host-genetics, microbiome and environmental factors at the mechanistic level will allow in the future to make better diagnostics and therefore specifically manipulate the microbiome with the aim to find personalized solutions for IBD, obesity, medication and the prediction of risk for chronic inflammation.

5 References

1. C. J. Conselice, A. Wilkinson, K. Duncan, A. Mortlock, The evolution of galaxy number density at $z < 8$ and its implications. *Astrophys. J.* **830**, 83 (2016).
2. A. J. Macpherson, K. D. McCoy, Standardised animal models of host microbial mutualism. *Mucosal Immunol.* **8**, 476–486 (2015).
3. C. Huttenhower *et al.*, Structure, function and diversity of the healthy human microbiome. *Nature*. **486**, 207–214 (2012).
4. R. Sender, S. Fuchs, R. Milo, Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, e1002533 (2016).
5. G. L. Simon, S. L. Gorbach, Intestinal flora in health and disease. *Gastroenterology*. **86**, 174–193 (1984).
6. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2015).
7. W. S. Garrett, J. I. Gordon, L. H. Glimcher, Homeostasis and Inflammation in the Intestine. *Cell*. **140**, 859–870 (2010).
8. J. Li *et al.*, An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* **32**, 834–841 (2014).
9. T. Bhattacharya, T. S. Ghosh, S. S. Mande, Global Profiling of Carbohydrate Active Enzymes in Human Gut Microbiome. *PLoS One*. **10**, e0142038 (2015).
10. E. C. Martens, N. M. Koropatkin, T. J. Smith, J. I. Gordon, Complex Glycan Catabolism by the Human Gut Microbiota: The Bacteroidetes Sus-like Paradigm. *J. Biol. Chem.* **284**, 24673–24677 (2009).
11. C. A. Lozupone, R. Knight, Species divergence and the measurement of microbial diversity. *FEMS Microbiol. Rev.* **32**, 557–78 (2008).
12. R. E. Ley, C. A. Lozupone, M. Hamady, R. Knight, J. I. Gordon, Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6**, 776–88 (2008).
13. I. I. Ivanov, K. Honda, Intestinal commensal microbes as immune modulators. *Cell Host Microbe*. **12**, 496–508 (2012).
14. L. V. Hooper, D. R. Littman, A. J. Macpherson, Interactions Between the Microbiota and the Immune System. *Science (80-.).* **336**, 1268–1273 (2012).
15. S. V. Lynch, O. Pedersen, The Human Intestinal Microbiome in Health and Disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).

16. J. L. Round, S. K. Mazmanian, The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* **9**, 313–323 (2009).
17. Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome. *Nature*. **486**, 207–214 (2012).
18. M. Arumugam *et al.*, Enterotypes of the human gut microbiome. *Nature*. **473**, 174–180 (2011).
19. T. Ding, P. D. Schloss, Dynamics and associations of microbial community types across the human body. *Nature*. **509**, 357–360 (2014).
20. M. J. Claesson *et al.*, Gut microbiota composition correlates with diet and health in the elderly. *Nature*. **488**, 178–184 (2012).
21. A. Gorvitovskaia, S. P. Holmes, S. M. Huse, Interpreting *Prevotella* and Bacteroides as biomarkers of diet and lifestyle. *Microbiome*. **4**, 15 (2016).
22. P. I. Costea *et al.*, Enterotypes in the landscape of gut microbial community composition. *Nat. Microbiol.* **3**, 8–16 (2018).
23. L. R. Lopetuso, F. Scaldaferri, V. Petito, A. Gasbarrini, Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog.* **5**, 23 (2013).
24. S. R. Gill *et al.*, Metagenomic analysis of the human distal gut microbiome. *Science*. **312**, 1355–9 (2006).
25. P. B. Eckburg *et al.*, Diversity of the Human Intestinal Microbial Flora. *Science (80-.).* **308**, 1635–1638 (2005).
26. K. Atarashi *et al.*, Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature*. **500**, 232–236 (2013).
27. I. I. Ivanov *et al.*, Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*. **139**, 485–498 (2009).
28. M. D. COLLINS *et al.*, The Phylogeny of the Genus Clostridium: Proposal of Five New Genera and Eleven New Species Combinations. *Int. J. Syst. Bacteriol.* **44**, 812–826 (1994).
29. N. Yutin, M. Y. Galperin, A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ. Microbiol.* **15**, 2631–41 (2013).
30. C. D. Garland, A. Lee, M. R. Dickson, Segmented filamentous bacteria in the rodent small intestine: Their colonization of growing animals and possible role in host resistance toSalmonella. *Microb. Ecol.* **8**, 181–190 (1982).
31. P. M. Smith *et al.*, The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis. *Science (80-.).* **341**, 569–573 (2013).

32. N. Arpaia *et al.*, Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature*. **504**, 451–5 (2013).
33. H. M. Wexler, Bacteroides: the good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
34. A. G. Wexler, A. L. Goodman, An insider's perspective: Bacteroides as a window into the microbiome. *Nat. Microbiol.* **2**, 17026 (2017).
35. J. Xu *et al.*, A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. *Science (80-.).* **299**, 2074–2076 (2003).
36. E. C. Martens *et al.*, Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *PLoS Biol.* **9**, e1001221 (2011).
37. S. M. Lee *et al.*, Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature*. **501**, 426–9 (2013).
38. S. K. Mazmanian, C. H. Liu, A. O. Tzianabos, D. L. Kasper, An Immunomodulatory Molecule of Symbiotic Bacteria Directs Maturation of the Host Immune System. *Cell*. **122**, 107–118 (2005).
39. M. M. Curtis *et al.*, The gut commensal Bacteroides thetaiotaomicron exacerbates enteric infection through modification of the metabolic landscape. *Cell Host Microbe*. **16**, 759–69 (2014).
40. J. M. Larsen, The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology*. **151**, 363 (2017).
41. H. Hayashi, K. Shibata, M. Sakamoto, S. Tomita, Y. Benno, *Prevotella* copri sp. nov. and *Prevotella* stercorea sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **57**, 941–6 (2007).
42. S. A. Smits *et al.*, Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science (80-.).* **357**, 802–806 (2017).
43. J. C. Clemente *et al.*, The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183–e1500183 (2015).
44. I. Martínez *et al.*, The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep.* **11**, 527–538 (2015).
45. P. Kovatcheva-Datchary *et al.*, Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of *Prevotella*. *Cell Metab.* **22**, 971–982 (2015).
46. J. U. Scher *et al.*, Expansion of intestinal *Prevotella* copri correlates with enhanced susceptibility to arthritis. *Elife*. **2**, e01202 (2013).

47. E. Elinav *et al.*, NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*. **145**, 745–57 (2011).
48. A. A. Fodor *et al.*, The “Most Wanted” Taxa from the Human Microbiome for Whole Genome Sequencing. *PLoS One*. **7**, e41294 (2012).
49. J. K. Goodrich *et al.*, Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host Microbe*. **19**, 731–743 (2016).
50. J. K. Goodrich *et al.*, Human Genetics Shape the Gut Microbiome. *Cell*. **159**, 789–799 (2014).
51. H. Xie *et al.*, Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. *Cell Syst.* **3**, 572–584.e3 (2016).
52. J. Wang *et al.*, Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nat. Genet.* **48**, 1396–1406 (2016).
53. L. A. David *et al.*, Diet rapidly and reproducibly alters the human gut microbiome. *Nature*. **505**, 559–563 (2014).
54. R. N. Carmody *et al.*, Diet Dominates Host Genotype in Shaping the Murine Gut Microbiota. *Cell Host Microbe*. **17**, 72–84 (2015).
55. A. Cotillard *et al.*, Dietary intervention impact on gut microbial gene richness. *Nature*. **500**, 585–588 (2013).
56. L. Maier *et al.*, Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*. **555**, 623–628 (2018).
57. T. Yatsunenko *et al.*, Human gut microbiome viewed across age and geography. *Nature*. **486**, 222–227 (2012).
58. Y. He *et al.*, Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* (2018), doi:10.1038/s41591-018-0164-x.
59. A. Salonen, W. M. de Vos, Impact of Diet on Human Intestinal Microbiota and Health. *Annu. Rev. Food Sci. Technol.* **5**, 239–262 (2014).
60. J. Aron-Wisnewsky, K. Clément, The gut microbiome, diet, and links to cardiometabolic and chronic disorders. *Nat. Rev. Nephrol.* **12**, 169–81 (2016).
61. A. Rogowski *et al.*, Glycan complexity dictates microbial resource allocation in the large intestine. *Nat. Commun.* **6**, 7481 (2015).
62. M. Stehr *et al.*, Charles River altered Schaedler flora (CRASF®) remained stable for four years in a mouse colony housed in individually ventilated cages. *Lab. Anim.* **43**, 362–370 (2009).

63. E. J. C. Gálvez, A. Iljazovic, A. Gronow, R. Flavell, T. Strowig, Shaping of Intestinal Microbiota in Nlrp6- and Rag2-Deficient Mice Depends on Community Structure. *Cell Rep.* **21** (2017), doi:10.1016/j.celrep.2017.12.027.
64. E. J. C. Gálvez, A. Iljazovic, A. Gronow, R. Flavell, T. Strowig, Shaping of Intestinal Microbiota in Nlrp6- and Rag2-Deficient Mice Depends on Community Structure. *Cell Rep.* **21**, 3914–3926 (2017).
65. U. Roy *et al.*, Distinct Microbial Communities Trigger Colitis Development upon Intestinal Barrier Damage via Innate or Adaptive Immune Cells. *Cell Rep.* **21**, 994–1008 (2017).
66. P. J. Turnbaugh *et al.*, A core gut microbiome in obese and lean twins. *Nature*. **457**, 480–484 (2009).
67. A. L. Goodman *et al.*, Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl. Acad. Sci.* **108**, 6252–6257 (2011).
68. T. Matsuki, K. Watanabe, J. Fujimoto, T. Takada, R. Tanaka, Use of 16S rRNA Gene-Targeted Group-Specific Primers for Real-Time PCR Analysis of Predominant Bacteria in Human Feces. *Appl. Environ. Microbiol.* **70**, 7220–7228 (2004).
69. J. G. Caporaso *et al.*, Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* **108**, 4516–4522 (2011).
70. A. Bankevich *et al.*, SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
71. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. **27**, 578–579 (2011).
72. R. K. Aziz *et al.*, The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*. **9**, 75 (2008).
73. R. Overbeek *et al.*, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2014).
74. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15–21 (2013).
75. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
76. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

77. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
78. J. Huerta-Cepas *et al.*, eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44**, D286-93 (2016).
79. J. P. Meier-Kolthoff, A. F. Auch, H.-P. Klenk, M. Göker, Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. **14**, 60 (2013).
80. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
81. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–3 (2014).
82. I. Lagkouvardos *et al.*, IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* **6**, 33721 (2016).
83. N. Terrapon, V. Lombard, H. J. Gilbert, B. Henrissat, Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics*. **31**, 647–55 (2015).
84. H. Zhang *et al.*, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
85. P. J. McMurdie, S. Holmes, phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One*. **8**, e61217 (2013).
86. H. Wickham, *ggplot2: elegant graphics for data analysis* (Springer New York, 2009).
87. N. Segata *et al.*, Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
88. P. J. McMurdie *et al.*, Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).
89. T. S. Stappenbeck, H. W. Virgin, Accounting for reciprocal host?microbiome interactions in experimental science. *Nature*. **534**, 191–199 (2016).
90. J. Henao-Mejia *et al.*, Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature*. **482**, 179 (2012).
91. B. Hu *et al.*, Microbiota-induced activation of epithelial IL-6 signaling links inflammasome-driven inflammation with transmissible cancer. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9862–9867 (2013).

92. A. Rehman *et al.*, Nod2 is essential for temporal development of intestinal microbial communities. *Gut*. **60**, 1354–1362 (2011).
93. M. Vijay-Kumar *et al.*, Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science*. **328**, 228–31 (2010).
94. H. Zhang, J. B. Sparks, S. V Karyala, R. Settlage, X. M. Luo, Host adaptive immunity alters gut microbiota. *ISME J.* **9**, 770–781 (2015).
95. M. Mamantopoulos *et al.*, Nlrp6- and ASC-Dependent Inflammasomes Do Not Shape the Commensal Gut Microbiota Composition. *Immunity*. **47**, 339–348.e4 (2017).
96. S. J. Robertson *et al.*, Nod1 and Nod2 signaling does not alter the composition of intestinal bacterial communities at homeostasis. *Gut Microbes*. **4**, 222–31 (2013).
97. C. Ubeda *et al.*, Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J. Exp. Med.* **209**, 1445–1456 (2012).
98. P. Lemire *et al.*, The NLR Protein NLRP6 Does Not Impact Gut Microbiota Composition. *Cell Rep.* **21**, 3653–3661 (2017).
99. W. S. Garrett *et al.*, Communicable Ulcerative Colitis Induced by T-bet Deficiency in the Innate Immune System. *Cell*. **131**, 33–45 (2007).
100. S. Kawamoto *et al.*, Foxp3+ T Cells Regulate Immunoglobulin A Selection and Facilitate Diversification of Bacterial Species Responsible for Immune Homeostasis. *Immunity*. **41**, 152–165 (2014).
101. T. S. Stappenbeck, H. W. Virgin, Accounting for reciprocal host–microbiome interactions in experimental science. *Nature*. **534**, 191–199 (2016).
102. P. Rausch *et al.*, Analysis of factors contributing to variation in the C57BL/6J fecal microbiota across German animal facilities. *Int. J. Med. Microbiol.* **306**, 343–355 (2016).
103. A. J. Macpherson, K. D. McCoy, Standardised animal models of host microbial mutualism. *Mucosal Immunol.* **8**, 476–486 (2014).
104. N. W. Palm *et al.*, Immunoglobulin A Coating Identifies Colitogenic Bacteria in Inflammatory Bowel Disease. *Cell*. **158**, 1000–1010 (2014).
105. G. F. Sonnenberg *et al.*, Innate lymphoid cells promote anatomical containment of lymphoid-resident commensal bacteria. *Science*. **336**, 1321–5 (2012).
106. M. Levy *et al.*, Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating NLRP6 Inflammasome Signaling. *Cell*. **163**, 1428–1443 (2015).

107. M. Wlodarska *et al.*, NLRP6 inflammasome orchestrates the colonic host-microbial interface by regulating goblet cell mucus secretion. *Cell*. **156**, 1045–1059 (2014).
108. D. Berry *et al.*, Phylotype-level 16S rRNA analysis reveals new bacterial indicators of health state in acute murine colitis. *ISME J.* **6**, 2091–106 (2012).
109. A. Rodríguez-Blanco, J.-F. Ghiglione, P. Catala, E. O. Casamayor, P. Lebaron, Spatial comparison of total vs. active bacterial populations by coupling genetic fingerprinting and clone library analyses in the NW Mediterranean Sea. *FEMS Microbiol. Ecol.* **67**, 30–42 (2009).
110. J. B. Emerson *et al.*, Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome*. **5**, 86 (2017).
111. M. Arumugam *et al.*, Enterotypes of the human gut microbiome. *Nature*. **473**, 174–180 (2011).
112. P. J. Turnbaugh *et al.*, Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc. Natl. Acad. Sci.* **107**, 7503–7508 (2010).
113. E. A. Franzosa *et al.*, Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci.* **111**, E2329–E2338 (2014).
114. A. Heintz-Buschart, P. Wilmes, Human Gut Microbiome: Function Matters. *Trends Microbiol.* (2017), doi:10.1016/j.tim.2017.11.002.
115. M. Wu *et al.*, Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. *Science (80-.).* **350**, aac5992-aac5992 (2015).
116. B. Lim, M. Zimmermann, N. A. Barry, A. L. Goodman, Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut. *Cell*. **169**, 547–558.e15 (2017).
117. P. Joglekar *et al.*, Genetic Variation of the SusC/SusD Homologs from a Polysaccharide Utilization Locus Underlies Divergent Fructan Specificities and Functional Adaptation in Bacteroides thetaiotaomicron Strains. *mSphere*. **3** (2018), doi:10.1128/mSphereDirect.00185-18.
118. D. Dodd, Y.-H. Moon, K. Swaminathan, R. I. Mackie, I. K. O. Cann, Transcriptomic analyses of xylan degradation by *Prevotella* bryantii and insights into energy acquisition by xylanolytic bacteroidetes. *J. Biol. Chem.* **285**, 30261–73 (2010).
119. H. Hayashi, K. Shibata, M. Sakamoto, S. Tomita, Y. Benno, *Prevotella* copri sp. nov. and *Prevotella* stercorea sp. nov., isolated from human faeces. *Int. J. Syst.*

Evol. Microbiol. **57**, 941–946 (2007).

120. P. Kovatcheva-Datchary *et al.*, Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of *Prevotella*. *Cell Metab.* **22**, 971–982 (2015).
121. J. U. Scher *et al.*, Expansion of intestinal *Prevotella* copri correlates with enhanced susceptibility to arthritis. *Elife* (2013), doi:10.7554/eLife.01202.
122. E. Elinav *et al.*, NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*. **145**, 745–57 (2011).
123. R. E. Ley, *Prevotella* in the gut: choose carefully. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 69–70 (2016).
124. T. Chen *et al.*, Fiber-utilizing capacity varies in *Prevotella*- versus Bacteroides-dominated gut microbiota. *Sci. Rep.* **7**, 2594 (2017).
125. E. J. C. Gálvez, A. Iljazovic, A. Gronow, R. Flavell, T. Strowig, Shaping of Intestinal Microbiota in Nlrp6- and Rag2-Deficient Mice Depends on Community Structure. *Cell Rep.* **21**, 3914–3926 (2017).
126. S. Thiemann *et al.*, Enhancement of IFNγ Production by Distinct Commensals Ameliorates Salmonella -Induced Disease. *Cell Host Microbe*. **21**, 682–694.e5 (2017).
127. T. Clavel, I. Lagkouvardos, M. Blaut, B. Stecher, The mouse gut microbiome revisited: From complex diversity to model ecosystems. *Int. J. Med. Microbiol.* **306**, 316–327 (2016).
128. P. Rausch *et al.*, Analysis of factors contributing to variation in the C57BL/6J fecal microbiota across German animal facilities. *Int. J. Med. Microbiol.* **306**, 343–55 (2016).
129. R. E. Ley *et al.*, Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci.* **102**, 11070–11075 (2005).
130. V. Gupta, N. M. Chaudhari, S. Iskepalli, C. Dutta, Divergences in gene repertoire among the reference *Prevotella* genomes derived from distinct body sites of human. *BMC Genomics*. **16**, 153 (2015).
131. F. De Filippis, N. Pellegrini, L. Laghi, M. Gobbetti, D. Ercolini, Unusual sub-genus associations of faecal *Prevotella* and Bacteroides with specific dietary patterns. *Microbiome*. **4**, 57 (2016).
132. S. Rakoff-Nahoum, K. R. Foster, L. E. Comstock, The evolution of cooperation within the gut microbiota. *Nature*. **533**, 255–259 (2016).
133. Y. Wang, N. B. Shoemaker, A. A. Salyers, Regulation of a Bacteroides operon

that controls excision and transfer of the conjugative transposon CTnDOT. *J. Bacteriol.* **186**, 2548–57 (2004).

134. N. M. Koropatkin, E. A. Cameron, E. C. Martens, How glycan metabolism shapes the human gut microbiota. *Nat. Rev. Microbiol.* **10**, 323–35 (2012).
135. T. S. Stappenbeck, H. W. Virgin, Accounting for reciprocal host-microbiome interactions in experimental science. *Nature*. **534**, 191–9 (2016).
136. R. Orcutt, RP; Gianni, FJ; Judge, Development of an “altered Schaedler flora” for NCI gnotobiotic rodents. *Microecol. Ther.*, 59 (1987).
137. P. Trexler, R. Orcutt, *Development of gnotobiotics and contamination control in laboratory animal science* (Memphis: American Association for Laboratory Animal Science, 1999).
138. S. Brugiroux *et al.*, Genome-guided design of a defined mouse microbiota that confers colonization resistance against Salmonella enterica serovar Typhimurium. *Nat. Microbiol.* **2**, 16215 (2017).
139. M. Wymore Brand *et al.*, The Altered Schaedler Flora: Continued Applications of a Defined Murine Microbial Community. *ILAR J.* **56**, 169–178 (2015).
140. E. Norin, T. Midtvedt, Intestinal microflora functions in laboratory mice claimed to harbor a "normal" intestinal microflora. Is the SPF concept running out of date? *Anaerobe*. **16**, 311–3 (2010).
141. A. J. Błażejewski *et al.*, Microbiota Normalization Reveals that Canonical Caspase-1 Activation Exacerbates Chemically Induced Intestinal Inflammation. *Cell Rep.* **19** (2017), doi:10.1016/j.celrep.2017.05.058.
142. U. Roy *et al.*, Distinct Microbial Communities Trigger Colitis Development upon Intestinal Barrier Damage via Innate or Adaptive Immune Cells. *Cell Rep.* **21**, 994–1008 (2017).
143. S. Thiemann *et al.*, Enhancement of IFNγ Production by Distinct Commensals Ameliorates Salmonella-Induced Disease. *Cell Host Microbe*. **21** (2017), doi:10.1016/j.chom.2017.05.005.
144. S. S. Seregin *et al.*, NLRP6 Protects Il10 −/− Mice from Colitis by Limiting Colonization of Akkermansia muciniphila. *Cell Rep.* **19**, 733–745 (2017).
145. M. Levy, H. Shapiro, C. A. Thaiss, E. Elinav, NLRP6: A Multifaceted Innate Immune Sensor. *Trends Immunol.* **38**, 248–260 (2017).
146. K. D. McCoy, M. B. Geuking, F. Ronchi, *Curr. Protoc. Immunol.*, in press, doi:10.1002/cpim.25.
147. D. Rothschild *et al.*, Environment dominates over host genetics in shaping human

gut microbiota. *Nature*. **555**, 210–215 (2018).

148. G. D. Wu *et al.*, Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science (80-.).* **334**, 105–108 (2011).
149. S. L. Schnorr *et al.*, Gut microbiome of the Hadza hunter-gatherers. *Nat. Commun.* **5**, 3654 (2014).
150. J. C. Clemente *et al.*, The microbiome of uncontacted Amerindians. *Sci. Adv.* **1**, e1500183–e1500183 (2015).
151. I. Martínez *et al.*, The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep.* **11**, 527–538 (2015).
152. E. Elinav *et al.*, NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell*. **145**, 745–757 (2011).
153. R. E. Ley, Gut microbiota in 2015: *Prevotella* in the gut: choose carefully. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 69–70 (2016).
154. M. Zhang *et al.*, Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3708-17 (2014).
155. D. Dodd, Y.-H. Moon, K. Swaminathan, R. I. Mackie, I. K. O. Cann, Transcriptomic Analyses of Xylan Degradation by *Prevotella bryantii* and Insights into Energy Acquisition by Xylanolytic Bacteroidetes. *J. Biol. Chem.* **285**, 30261–30273 (2010).
156. Z. A. Costliow, P. H. Degnan, Thiamine Acquisition Strategies Impact Metabolism and Competition in the Gut Microbe Bacteroides thetaiotaomicron. *mSystems*. **2**, e00116-17 (2017).
157. M. J. Merrick, R. A. Edwards, Nitrogen control in bacteria. *Microbiol. Rev.* **59**, 604–22 (1995).
158. K. A. PITTMAN, M. P. BRYANT, PEPTIDES AND OTHER NITROGEN SOURCES FOR GROWTH OF BACTEROIDES RUMINICOLA. *J. Bacteriol.* **88**, 401–10 (1964).
159. J. N. Kim *et al.*, Metabolic networks for nitrogen utilization in *Prevotella* ruminicola 23. *Sci. Rep.* **7**, 7851 (2017).
160. A. Birkett, J. Muir, J. Phillips, G. Jones, K. O'Dea, Resistant starch lowers fecal concentrations of ammonia and phenols in humans. *Am. J. Clin. Nutr.* **63**, 766–772 (1996).
161. F. E. Rey *et al.*, Metabolic niche of a prominent sulfate-reducing human gut

bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 13582–7 (2013).

162. H. Li *et al.*, The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6**, 8292 (2015).
163. A. A. Salyers, N. B. Shoemaker, L. Y. Li, In the driver's seat: the Bacteroides conjugative transposons and the elements they mobilize. *J. Bacteriol.* **177**, 5727–31 (1995).
164. B. R. tenOever, The Evolution of Antiviral Defense Systems. *Cell Host Microbe*. **19**, 142–149 (2016).
165. A. G. Wexler, A. L. Goodman, An insider's perspective: Bacteroides as a window into the microbiome. *Nat. Microbiol.* **2**, 17026 (2017).
166. G. D. Wu *et al.*, Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. *Science (80-.).* **334**, 105–108 (2011).
167. B. Nograšek, T. Accetto, L. Fanedl, G. Avguštin, Description of a novel pectin-degrading bacterial species *Prevotella* pectinovora sp. nov., based on its phenotypic and genomic traits. *J. Microbiol.* **53**, 503–510 (2015).
168. D. Ndeh *et al.*, Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature*. **544**, 65–70 (2017).

Declaration of Experimental Assistance

Results section 3.1

I performed the experimental design and analyzed all the data presented in this section. Aida Iljazovic and Achim Gronow supported the experiments by handling of mice, fecal transplantation and sampling of the experiment as presented in the Figure 3.1.7, panel A. Samples from this experiment were processed and analyzed by me as presented in Figures 3.1.7 to 3.1.13.

Results section 3.2

I designed and analyzed all the data presented in this section. Aida Iljazovic contributed with the technical support by handling of mice and *Prevotella* colonization presented in the figures 3.2.5-A, figures 3.2.6-A, figures 3.2.7-A and figures 3.2.9-A,D.